

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
22 February 2001 (22.02.2001)

PCT

(10) International Publication Number  
WO 01/12817 A1

- (51) International Patent Classification<sup>7</sup>: C12N 15/52, C12P 1/00, 19/62
- (21) International Application Number: PCT/US00/22080
- (22) International Filing Date: 11 August 2000 (11.08.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:  
60/148,848 12 August 1999 (12.08.1999) US  
09/637,309 11 August 2000 (11.08.2000) US
- (71) Applicants (for all designated States except US): MAXY-GEN, INC. [US/US]; 515 Galveston Drive, Redwood City, CA 94063 (US). LIU, Lu [CN/US]; 519 Skiff Circle, Redwood City, CA 94065 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): KREBBER, Claus [DE/US]; 1935 Rock Street #1, Mountain View, CA 94043 (US). DAVIS, S., Christopher [GB/US]; 118 Church Street #103, San Francisco, CA 94114 (US). DELCARDAYRE, Stephen [US/US]; 2049 Monroe Avenue, Belmont, CA 94002 (US). SELIFONOV, Sergey, A. [RU/US]; 2240 Homestead Court, Los Altos, CA 94024 (US). HOWARD, Russell [AU/US]; 12700 Viscaino Road, Los Altos Hills, CA 94022 (US).
- (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.
- (84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:  
— With international search report.  
— Before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments.
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 01/12817 A1

(54) Title: EVOLUTION AND USE OF ENZYMES FOR COMBINATORIAL AND MEDICINAL CHEMISTRY

(57) Abstract: This invention provides libraries of recombinant derivatizing enzymes that are useful for biocatalytic synthesis of derivatives of organic molecules, including lead compounds for pharmaceutical use. The recombinant derivatizing enzymes catalyze reactions such as modification or replacement of functional groups on the organic molecules, or addition of chemical moieties onto preexisting functional groups. The use of recombinant enzyme libraries enables one to obtain enzymes that catalyze the formation of organic molecule derivatives that could not otherwise be made using only naturally occurring enzymes.

## EVOLUTION AND USE OF ENZYMES FOR COMBINATORIAL AND MEDICINAL CHEMISTRY

5

This application claims the benefit of U.S. Provisional Application No. 60/148,848, filed August 12, 1999, the entire disclosure of which is hereby incorporated by reference.

### COPYRIGHT NOTIFICATION PURSUANT TO 37 C.F.R. § 1.71(e)

10

A portion of this disclosure contains material which is subject to copyright protection. The copyright owner has no objection to the facsimile reproduction by anyone of the patent document or the patent disclosure, as it appears in the Patent and Trademark Office patent file or records, but otherwise reserves all copyright rights whatsoever.

### BACKGROUND OF THE INVENTION

#### 15 Field of the Invention

This invention pertains to the field of enzymatic synthesis of combinatorial libraries of organic molecules using evolved enzymes. The invention provides libraries of enzymes that, through directed evolution, are capable of biocatalytically synthesizing a multitude of derivatives of organic molecules. The libraries of organic molecule derivatives  
20 can be screened to identify active compounds, such as antibiotics and other therapeutic reagents, herbicides and pesticides, and the like.

#### Background

In the process of drug discovery, optimization of a lead compound represents one of many challenges. Very often, the lead compound lacks some of the pharmacological  
25 properties required for a fully functional pharmaceutical, such as high potency, selectivity, low toxicity, bioavailability, and the like. Additional modification of the lead compound is therefore often necessary for achieving an optimized drug that has a complete combination of desired properties. The traditional approach to derivatization depends upon a large body

of empirical experience to guide the medicinal chemist in the choice of which chemical analogs to synthesize and test. Some compounds are chosen for synthesis, and others are not. Similarly, when combinatorial chemistry is used to generate derivatives of lead compounds, particular building blocks are chosen for parallel synthesis of many analogs. Other building blocks are not. These choices are generally made in accordance with the body of experience in medicinal chemistry which can provide guidance as to those modifications that are likely to result in improvements, and those modifications that are likely to result in new undesired properties or exacerbation of existing properties. Unfortunately, however, this body of experience is for the most part specific to the individual medicinal chemist as it is not fully described, except in fragmentary form in numerous volumes.

Improvement of lead compounds having potential for pharmaceutical use is not the only situation in which derivatization of organic molecules is of interest. Organic molecules have many uses, including, for example, pesticides, herbicides, and others. To obtain compounds that exhibit improved properties for a particular application, it is often desirable to generate libraries of organic molecule derivatives that can then be screened to identify those derivatives that exhibit the desired properties.

Combinatorial synthesis methods have the potential to provide a way to synthesize a wide variety of lead compound derivatives without the need for *a priori* assumptions as to which derivatives are likely to be most favorable. Instead of synthesizing derivatives individually and testing them, one can make a large number of different derivatives simultaneously. Combinatorial synthesis is useful not only for the derivatization of lead compounds, but also for the synthesis of compounds that are screened to identify those that are worthy of further study as potential lead compounds. However, synthesis of combinatorial libraries of organic molecule derivatives is severely limited because many types of derivatives of organic molecules are difficult or even impossible to synthesize by purely chemical means.

Enzymes provide a potentially attractive route to the synthesis of chemical compound libraries from which one can identify those compounds that exhibit desired properties. Enzymes can act on mixtures of complex molecules in solution, catalyzing the synthesis of derivatives of the molecules without the production of byproducts. While traditional chemical processes for lead compound derivatization are typically non-selective

and require multiple protection and de-protection steps, such steps are not required for enzymatic synthesis. Moreover, enzymes can function under relatively mild conditions that are not destructive to the reaction products. Furthermore, enzymes can carry out several different types of modifications to organic molecules, such as existing and potential lead  
5 compounds and other biologically active molecules of interest. For example, enzymes can catalyze the addition of a moiety to a compound (*e.g.*, by ester, amide, carbonate, carbamate or glycoside linkage, and the like). Enzymes can also add new functional groups to an organic molecule, or can modify existing functional groups that are present on the compound. Enzymatic biocatalysis can also provide certain further advantages such as  
10 substrate-, stereo- and regio-selectivity.

Although enzymatic combinatorial biocatalysis has great potential, significant drawbacks remain. For example, a sufficiently wide variety of enzymes that can facilitate a full range of organic molecule derivatizations is not yet available. It is unlikely that one could obtain, from a set of naturally occurring enzymes, an enzyme that will possess the  
15 desired substrate, stereo- or regio- specificity for any particular organic molecule of interest. Thus, a need exists for derivatizing enzymes that are capable of producing a wide variety of organic molecule derivatives. The absence of such enzymes limits the number and type of organic molecule derivatives that are obtainable by combinatorial biocatalysis. Thus, a need exists for methods to obtain derivatizing enzymes that catalyze a wide variety of different  
20 organic molecule derivatizations, as well as for libraries of such organic molecule derivatives. The present invention fulfills these and other needs.

### SUMMARY OF THE INVENTION

The present invention provides methods for obtaining a library of organic molecule derivatives. The methods involve contacting an organic molecule with one or more  
25 members of a library of recombinant derivatizing enzymes and other necessary reactants to form the library of organic molecule derivatives. The derivatizing enzymes catalyze a reaction such as: a) modification of one or more functional groups present on the organic molecule; b) addition of a chemical moiety onto one or more functional groups present on the organic molecule; or c) introduction of a new functional group onto the organic  
30 molecule. The methods are useful for a wide variety of organic molecules, including, for



example, those that have pharmacological, herbicide, pesticide, or other activities, or are useful in industrial processes.

In some embodiments, the methods further involve performing one or more additional reactions on the derivatives that are obtained by contact with the derivatizing enzymes. Thus, the products of the initial reaction serve as intermediates for further reactions. The further reactions can involve, for example, contacting the library of organic molecule derivatives with one or more members of a second library of recombinant derivatizing enzymes and other necessary reactants to form a further library of organic molecule derivatives. Alternatively, the intermediates can be modified chemically or with other enzymes.

The libraries of recombinant derivatizing enzymes are obtained, in some embodiments, by (1) recombining at least first and second forms of a nucleic acid that encodes a derivatizing enzyme, wherein the first and second forms differ from each other in two or more nucleotides, to produce a library of recombinant polynucleotides; and (2) expressing the library of recombinant polynucleotides to obtain the library of recombinant derivatizing enzymes. If desired, the method can further involve (3) recombining at least one recombinant polynucleotide that encodes a member of the library of recombinant derivatizing enzymes with a further form of the nucleic acid that encodes a derivatizing enzyme, which is the same or different from the first and second forms, to produce a further library of recombinant nucleic acids; (4) expressing the further library of recombinant polynucleotides to obtain a further library of recombinant derivatizing enzymes; and (5) repeating (3) and (4), as necessary, until the further library of recombinant derivatizing enzymes contains a desired number of different recombinant derivatizing enzymes.

The invention also provides methods of obtaining an enzyme that catalyzes the synthesis of a desired organic molecule derivative. These methods involve contacting an organic molecule with members of a library of recombinant derivatizing enzymes and other necessary reactants to form a library of organic molecule derivatives; identifying the desired organic molecule derivative in the library of organic molecule derivatives; and identifying the member of the library of recombinant derivatizing enzymes that catalyzes the synthesis of the desired organic molecule derivative.

Also provided by the invention are libraries of recombinant derivatizing enzymes, wherein the recombinant derivatizing enzymes, when contacted with an organic molecule having one or more functional groups, catalyze a reaction such as: a) modification of one or more of the functional groups; b) addition of a chemical moiety onto one or more of the functional groups; or c) introduction of a new functional group.

In another embodiment, the invention provides libraries of organic molecule derivatives. The libraries are biocatalytically synthesized by contacting an organic molecule having one or more functional groups with a plurality of members of a library of recombinant derivatizing enzymes that catalyze a reaction such as: a) modification of one or more of the functional groups; b) addition of a chemical moiety onto one or more of the functional groups; or c) introduction of a new functional group.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 shows potential sugar attachment points on vancomycin hydrochloride.

Figure 2 shows potential sugar attachment points on somatostatin.

Figure 3 shows potential sugar attachment points on cholic acid.

Figure 4 shows potential sugar attachment points on L-thyroxine.

Figure 5 shows potential sugar attachment points on nogalamycin.

Figure 6 shows potential sugar attachment points on syringaldazine.

Figure 7 shows potential sugar attachment points on alcarubicin.

Figure 8 shows potential sugar attachment points on ritodrine hydrochloride.

Figure 9 shows potential sugar attachment points on rifamycin.

Figure 10 shows sugar attachment points on ristomycin sulfate. Five additional hydroxyls on the backbone are also shown (but not indicated by arrows); these constitute potential sugar attachment points.

Figure 11 shows a multi-step chemical methylation of erythromycin A and its analogs.

Figure 12 shows the reaction catalyzed by S-adenosylmethionine (SAM) dependent methyltransferases.

Figure 13 shows the specificity of O-methyltransferases that can be shuffled to obtain recombinant enzymes that have 6-OMTase activity using erythromycin and its analogs as substrates.

Figure 14 shows DNA and protein sequence similarity of the O-methyltransferases that are shuffled to obtain recombinant enzymes that have 6-OMTase activity using erythromycin and its analogs as substrates.

Figure 15 shows a microtiter plate high-throughput primary screen for the identification of methyltransferases that have novel specificity.

Figure 16 shows a schematic of the use of erythromycin A 6-O-methyltransferase for the biocatalytic synthesis of clarithromycin.

Figure 17 shows a secondary assay for a clarithromycin synthase. MS/MS detection of a 590/158 pair identifies methylation of the macrolide ring.

Figure 18 shows a further secondary assay for a clarithromycin synthase. Phenyl Boronate reacts specifically with cis diols at neutral pH. Only clarithromycin has the 11-12-cis diol that can react to give an 834.5 ion.

Figure 19 shows a map of the vector pCKZEBB.

## DETAILED DESCRIPTION

### Definitions

A "derivatizing enzyme" is an enzyme that can catalyze a reaction on an organic molecule. For example, a derivatizing enzyme can modify an existing functional group that is present on the molecule, add a chemical moiety onto a functional group, or add a new functional group to the organic molecule. The organic molecules can include both synthetic (including, *e.g.*, non-naturally occurring compounds such as halo-containing compounds and the like) and naturally occurring compounds.

A "recombinant derivatizing enzyme" is a non-naturally occurring derivatizing enzyme that differs in sequence from a naturally occurring derivatizing enzyme by at least one amino acid residue. Recombinant derivatizing enzymes include derivatizing enzymes that are composed of a plurality of blocks of amino acids, which blocks are not contiguous in a naturally occurring enzyme. The blocks are generally of random length. A recombinant derivatizing enzyme may be chimeric, thus having portions of its sequence

derived from the sequences of at least two different parental enzymes. A chimeric recombinant derivatizing enzyme is encoded by a chimeric gene that contains nucleic acid segments derived from at least two distinct parental genes or parental gene segments. A parental gene may optionally encode a derivatizing enzyme.

5           As used herein, the term "library" refers to a collection of diverse molecules, such as, for example, recombinant derivatizing enzymes and organic compound analogues. Libraries of the present invention have at least two distinct member molecules but can vary in size. Typically, invention libraries have at least about 5 distinct members, and more typically at least about 10 distinct member molecules. Larger libraries of the present  
10          invention typically have at least about 100 distinct member molecules, sometimes more than about 10,000, or even more than about 100,000. Very large libraries of the present invention can have more than about 1,000,000 members.

          A "functional group" refers to an atom or group of atoms that define the structure of a particular family of organic compounds and determines their properties.  
15          Functional groups include, for example, alkenes, alkynes, aromatics, halogens, hydroxyls, ethers, esters, aldehydes, ketones, carboxylic acids, amides, amines, and the like.

          A "lead compound" is a prototype compound that has a desired biological or pharmacological activity, but may have other characteristics that are undesirable. For example, the lead compound might be toxic, insoluble, have other biological activities, have  
20          less than optimal bioavailability (e.g., properties such as absorption, distribution, metabolism, and excretion (i.e., ADME), or less than optimal biological activity, etc.

          "Nucleic acid" refers to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages,  
25          which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, peptide-nucleic acids (PNAs), and the like.

30          Unless otherwise indicated, a particular nucleic acid sequence also implicitly encompasses conservatively modified variants thereof (*e.g.*, degenerate codon substitutions)

and complementary sequences, as well as the sequence explicitly indicated. The term “nucleic acid” is used interchangeably herein with “gene,” “cDNA,” “mRNA,” “oligonucleotide,” and “polynucleotide.”

The terms “polypeptide,” “peptide,” and “protein” are used interchangeably  
5 herein to refer to a polymer of amino acid residues. The terms apply to amino acid polymers in which one or more amino acid residue is an analog or mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers.

The term “amino acid” refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner  
10 similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline,  $\gamma$ -carboxyglutamate, and O-phosphoserine. Amino acid analogs refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, an  $\alpha$  carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R  
15 group (*e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium). Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. Amino acid mimetics refer to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that function in a manner similar to a naturally  
20 occurring amino acid.

Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

25 “Conservatively modified variants” applies to both amino acid and nucleic acid sequences. With respect to particular nucleic acid sequences, conservatively modified variants refer to those nucleic acids which encode identical or essentially identical amino acid sequences, or where the nucleic acid does not encode an amino acid sequence, to essentially identical sequences. Specifically, degenerate codon substitutions may be  
30 achieved by generating sequences in which the third position of one or more selected (or all) codons is substituted with mixed-base and/or deoxyinosine residues (Batzer *et al.*, *Nucleic*

*Acid Res.* 19:5081 (1991); Ohtsuka *et al.*, *J. Biol. Chem.* 260:2605-2608 (1985); Rossolini *et al.*, *Mol. Cell. Probes* 8:91-98 (1994)). Because of the degeneracy of the genetic code, a large number of functionally identical nucleic acids encode any given protein. For instance, the codons GCA, GCC, GCG and GCU all encode the amino acid alanine. Thus, at every position where an alanine is specified by a codon, the codon can be altered to any of the corresponding codons described without altering the encoded polypeptide. Such nucleic acid variations are "silent variations," which are one species of conservatively modified variations. Every nucleic acid sequence recited herein that encodes a polypeptide also describes every possible silent variation of the nucleic acid. One of skill will recognize that each codon in a nucleic acid (except AUG, which, along with GUG in some organisms, is ordinarily the only codon for methionine, and TGG, which is ordinarily the only codon for tryptophan) can be modified to yield a functionally identical molecule. Accordingly, each silent variation of a nucleic acid which encodes a polypeptide is implicit in each described sequence.

As to amino acid sequences, one of skill will recognize that individual substitutions, deletions or additions to a nucleic acid, peptide, polypeptide, or protein sequence which alter, add or delete a single amino acid or a small percentage of amino acids in the encoded sequence is a "conservatively modified variant" where the alteration results in the substitution of an amino acid with a chemically similar amino acid. Conservative substitution tables providing functionally similar amino acids are well known in the art. Such conservatively modified variants are in addition to and do not exclude polymorphic variants, interspecies homologs, and alleles of the invention.

The term "shuffling" is used herein to indicate recombination between non-identical sequences, in some embodiments shuffling may include crossover via homologous recombination or via non-homologous recombination, such as via cre/lox and/or flp/frt systems. Shuffling can be carried out by employing a variety of different formats, including, for example, in vitro and in vivo shuffling formats, in silico shuffling formats, shuffling formats that utilize either double-stranded or single-stranded templates, primer-based shuffling formats, nucleic acid fragmentation-based shuffling formats, oligonucleotide-mediated shuffling formats, all of which are based on recombination events between non-

identical sequences and are described in more detail or reference herein below, as well as other similar recombination-based formats.

### **Description of the Preferred Embodiments**

The present invention provides libraries of recombinant derivatizing enzymes  
5 that are useful for generating combinatorial libraries of chemical compounds, in particular organic molecules. Also provided are libraries of organic molecule derivatives that are obtained using the recombinant derivatizing enzyme libraries. The libraries of organic molecule derivatives are useful, for example, to identify those derivatives that have a desired biological activity and thus are suitable for testing as lead compounds, *e.g.*, for  
10 pharmaceutical or other use, and for creating combinatorial libraries of derivatives of a previously identified lead compound for testing for improved pharmacological or other parameters. The chemical compounds are often organic molecules, including synthetic molecules (including, for example, non-naturally occurring compounds) and natural products such as, for example, antibiotics.

15 The libraries of recombinant derivatizing enzymes provided by the invention provide several advantages over previously available methods for obtaining libraries of organic molecule derivatives. For example, the recombinant library will contain enzymes that exhibit catalytic properties that differ from one another in features such as catalytic rates and constants, stereo-, regio- and enantiomeric specificity, multiplicity of substrate  
20 selectivity, product inhibition, stability in a solvent used for biocatalytic synthesis, stability in chemical processes in general, and the like. The resulting multitude of different enzymes thus increases the number of different compounds that can be generated by biocatalytic reactions. When one enzyme is used for biocatalysis with a single organic molecule and a single chemical moiety donor, generally only one derivative is generated. In contrast, a  
25 multitude of recombinant enzymes is likely to include enzymes that can catalyze different reactions relative to the original enzyme, and thus are able to generate different products even starting with the same substrates as used with the original enzyme. Moreover, the use of enzymes for the synthesis of organic compounds of interest greatly facilitates scale-up of the synthetic reaction.

In presently preferred embodiments, DNA shuffling or other methods of recursive recombination are used to generate the libraries of recombinant enzymes. DNA shuffling has proven very effective at improving the level of known activity of a biocatalyst. An additional value of this technology lies in the ability to generate catalytic activities that were previously unknown among wild-type enzymes. Thus, this technology provides a reliable means of biocatalyst generation that decreases or even obviates the need to obtain naturally occurring biocatalysts for a targeted reaction. DNA shuffling of a family of related genes, for example, generates functionally diverse gene libraries with different physical properties that span a more complex sequence space than can be found in nature for a particular protein. Since the novel members of these enzyme libraries have never been under selective pressures in an organism, they are unbiased and can be screened for new activities that are rare or non-existent in natural samples. Thus, one can create diverse and complex enzyme libraries that catalyze a spectrum of important chemistries. For example, the enzymes can catalyze modifications of functional groups that are present on organic molecules, addition of chemical moieties onto functional groups (*e.g.*, acylation, glycosylation, and methylation), and introduction of new functional groups into the organic molecule (*e.g.*, introduction of hydroxyl groups by oxidation, double bonds by reduction, and the like). The enzyme libraries can be used directly to synthesize a multitude of products starting from substrate mixtures, or to synthesize a specific compound starting from a defined substrate set. Alternatively, single members of the library of recombinant enzymes can be used to synthesize mixtures of compounds by contacting the members with a mixture of substrates. In a further alternative embodiment of the present invention, each single member of the library of recombinant derivatizing enzymes can be tested with a defined substrate set to identify enzymes that have new and useful substrate selectivities or other useful features.

The organic molecule derivatives that are thus synthesized can then be screened to identify those that have a desired property, or can be further modified by one or more additional chemical or enzymatic reactions. One can also screen the enzyme libraries to identify those enzymes that have new and useful substrate selectivities or other desirable features, and use the enzymes to produce desired compounds.



The recombinant enzymes obtained using the methods of the invention can be used *in vitro*, or can be expressed by microbial cells that carry out the biocatalysis. In some embodiments, the microorganisms are modified to express one or more derivatizing enzymes for efficient biocatalytic manufacturing of the derivatized products. For example, the

5 microorganisms can include one or more recombinant polynucleotides that encode the improved acyltransferases, glycosyltransferases, oxidases, methyltransferases, or other biocatalytic enzymes, which are then expressed by the microbial cells. These polynucleotides can be introduced into organisms that naturally produce the starting substrate of interest. For example, a polynucleotide that encodes a recombinant derivatizing

10 enzyme can be introduced into an organism that naturally produces, or has been engineered to produce, a polyketide or other antibiotic. Thus, the recombinant polynucleotides that encode recombinant derivatizing enzymes of the invention are useful for the *in vivo* derivatization of organic compounds for which the backbones were previously prepared, for *in vivo* derivatization of organic compounds in the organism that biosynthesizes the

15 backbone of the organic molecule, and for *in vitro* use to derivatize a previously prepared organic molecule.

#### A. *Creation of Recombinant Libraries*

The invention involves, in some embodiments, creating recombinant libraries of polynucleotides that are then screened to identify those library members that encode an

20 enzyme or other polypeptide that exhibits a desired property, *e.g.*, enhanced enzymatic activity, stereospecificity, regiospecificity and enantiospecificity, reduced susceptibility to inhibitors, processing stability (*e.g.*, solvent stability, pH stability, thermal stability, etc.), and the like. The recombinant libraries can be created using any of various methods, including those described herein. For example, a variety of nucleic acid shuffling protocols

25 are available and fully described in the art. The following publications describe a variety of such procedures and/or methods which can be incorporated into such procedures, as well as other diversity generating protocols: Stemmer, et al., (1999) "Molecular breeding of viruses for targeting and other clinical properties. Tumor Targeting" 4:1-4; Nasset et al. (1999) "DNA Shuffling of subgenomic sequences of subtilisin" *Nature Biotechnology* 17:893-896;

30 Chang et al. (1999) "Evolution of a cytokine using DNA family shuffling" *Nature*

- Biotechnology* 17:793-797; Minshull and Stemmer (1999) "Protein evolution by molecular breeding" *Current Opinion in Chemical Biology* 3:284-290; Christians et al. (1999) "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling" *Nature Biotechnology* 17:259-264; Crameri et al. (1998) "DNA shuffling of a family of genes from diverse species accelerates directed evolution" *Nature* 391:288-291;
- 5 Crameri et al. (1997) "Molecular evolution of an arsenate detoxification pathway by DNA shuffling," *Nature Biotechnology* 15:436-438; Zhang et al. (1997) "Directed evolution of an effective fucosidase from a galactosidase by DNA shuffling and screening" *Proceedings of the National Academy of Sciences, U.S.A.* 94:4504-4509; Patten et al. (1997) "Applications of DNA Shuffling to Pharmaceuticals and Vaccines" *Current Opinion in Biotechnology*
- 10 8:724-733; Crameri et al. (1996) "Construction and evolution of antibody-phage libraries by DNA shuffling" *Nature Medicine* 2:100-103; Crameri et al. (1996) "Improved green fluorescent protein by molecular evolution using DNA shuffling" *Nature Biotechnology* 14:315-319; Gates et al. (1996) "Affinity selective isolation of ligands from peptide libraries through display on a lac repressor 'headpiece dimer'" *Journal of Molecular Biology* 255:373-
- 15 386; Stemmer (1996) "Sexual PCR and Assembly PCR" In: *The Encyclopedia of Molecular Biology*. VCH Publishers, New York. pp.447-457; Crameri and Stemmer (1995) "Combinatorial multiple cassette mutagenesis creates all the permutations of mutant and wildtype cassettes" *BioTechniques* 18:194-195; Stemmer et al., (1995) "Single-step
- 20 assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides" *Gene*, 164:49-53; Stemmer (1995) "The Evolution of Molecular Computation" *Science* 270: 1510; Stemmer (1995) "Searching Sequence Space" *Bio/Technology* 13:549-553; Stemmer (1994) "Rapid evolution of a protein in vitro by DNA shuffling" *Nature* 370:389-391; and Stemmer (1994) "DNA shuffling by random fragmentation and reassembly: In vitro
- 25 recombination for molecular evolution." *Proceedings of the National Academy of Sciences, U.S.A.* 91:10747-10751.

Additional details regarding DNA shuffling methods are found in U.S. Patents by the inventors and their co-workers, including: United States Patent 5,605,793 to Stemmer (February 25, 1997), "METHODS FOR IN VITRO RECOMBINATION;" United

30 States Patent 5,811,238 to Stemmer et al. (September 22, 1998) "METHODS FOR GENERATING POLYNUCLEOTIDES HAVING DESIRED CHARACTERISTICS BY

ITERATIVE SELECTION AND RECOMBINATION;" United States Patent 5,830,721 to Stemmer et al. (November 3, 1998), "DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASSEMBLY;" United States Patent 5,834,252 to Stemmer, et al. (November 10, 1998) "END-COMPLEMENTARY POLYMERASE REACTION,"  
5 and United States Patent 5,837,458 to Minshull, et al. (November 17, 1998), "METHODS AND COMPOSITIONS FOR CELLULAR AND METABOLIC ENGINEERING."

In addition, details and formats for DNA shuffling protocols are found in a variety of PCT and foreign patent application publications, including: Stemmer and Cramer, "DNA MUTAGENESIS BY RANDOM FRAGMENTATION AND REASSEMBLY" WO  
10 95/22625; Stemmer and Lipschutz "END COMPLEMENTARY POLYMERASE CHAIN REACTION" WO 96/33207; Stemmer and Cramer "METHODS FOR GENERATING POLYNUCLEOTIDES HAVING DESIRED CHARACTERISTICS BY ITERATIVE SELECTION AND RECOMBINATION" WO 97/0078; Minshull and Stemmer, "METHODS AND COMPOSITIONS FOR CELLULAR AND METABOLIC  
15 ENGINEERING" WO 97/35966; Punnonen et al. "TARGETING OF GENETIC VACCINE VECTORS" WO 99/41402; Punnonen et al. "ANTIGEN LIBRARY IMMUNIZATION" WO 99/41383; Punnonen et al. "GENETIC VACCINE VECTOR ENGINEERING" WO 99/41369; Punnonen et al. OPTIMIZATION OF IMMUNOMODULATORY PROPERTIES OF GENETIC VACCINES WO 99/41368; Stemmer and Cramer, "DNA MUTAGENESIS  
20 BY RANDOM FRAGMENTATION AND REASSEMBLY" EP 0934999; Stemmer "EVOLVING CELLULAR DNA UPTAKE BY RECURSIVE SEQUENCE RECOMBINATION" EP 0932670; Stemmer et al., "MODIFICATION OF VIRUS TROPISM AND HOST RANGE BY VIRAL GENOME SHUFFLING" WO 99/23107; Apt et al., "HUMAN PAPILLOMAVIRUS VECTORS" WO 99/21979; Del Cardayre et al.  
25 "EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE RECOMBINATION" WO 98/31837; Patten and Stemmer, "METHODS AND COMPOSITIONS FOR POLYPEPTIDE ENGINEERING" WO 98/27230; Stemmer et al., "METHODS FOR OPTIMIZATION OF GENE THERAPY BY RECURSIVE SEQUENCE SHUFFLING AND SELECTION" WO 98/31487; Arnold et al. "RECOMBINATION OF  
30 POLYNUCLEOTIDE SEQUENCES USING RANDOM OR DEFINED PRIMERS" WO 98/42832; Arnold et al. "METHOD FOR CREATING POLYNUCLEOTIDE AND

POLYPEPTIDE SEQUENCES" WO 9929902; Vind, "AN *In vitro* METHOD FOR CONSTRUCTION OF A DNA LIBRARY," WO 9841653; and Borchert et al., "METHOD FOR CONSTRUCTING A LIBRARY USING DNA SHUFFLING," WO 9841622.

Certain U.S. Applications provide additional details regarding DNA shuffling

5 and related techniques, as well as other diversity generating methods, including "SHUFFLING OF CODON ALTERED GENES" by Patten et al. filed September 29, 1998, (USSN 60/102,362), January 29, 1999 (USSN 60/117,729), and September 28, 1999, USSN09/407,800 (Attorney Docket Number 20-28520US/PCT); "EVOLUTION OF WHOLE CELLS AND ORGANISMS BY RECURSIVE SEQUENCE

10 RECOMBINATION", by del Cardayre et al. filed July 15, 1998 (USSN 09/166,188), and July 15, 1999 (USSN 09/354,922); "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION" by Crameri et al., filed February 5, 1999 (USSN 60/118,813) and filed June 24, 1999 (USSN 60/141,049) and filed September 28, 1999 (USSN 09/408,392, Attorney Docket Number 02-29620US); and "USE OF CODON-BASED

15 OLIGONUCLEOTIDE SYNTHESIS FOR SYNTHETIC SHUFFLING" by Welch et al., filed September 28, 1999 (USSN 09/408,393, Attorney Docket Number 02-010070US); "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS" by Selifonov and Stemmer, filed February 5, 1999 (USSN 60/118854) and USSN 09/416,375 filed October 12, 1999.

20 Shuffling formats that employ single stranded templates are described in "METHODS AND COMPOSITIONS FOR POLYPEPTIDE ENGINEERING," WO 9827230, by Patten et al.; "SINGLE-STRANDED NUCLEIC ACID TEMPLATE-MEDIATED RECOMBINATION AND NUCLEIC ACID FRAGMENT ISOLATION" by Affholter, USSN 60/186,482 filed March 2, 2000; "METHODS FOR GENERATING HIGHLY DIVERSE LIBRARIES," WO

25 0000632; and "METHOD FOR OBTAINING *IN VITRO* RECOMBINED POLYNUCLEOTIDE SEQUENCES, SEQUENCE BANKS, AND RESULTING SEQUENCES," WO 0009679.

As review of the foregoing publications, patents, published applications and U.S. patent applications reveals, shuffling of nucleic acids to provide new nucleic acids with

30 desired properties can be carried out by a number of established recombination methods and

these procedures can be combined with any of a variety of other diversity generating methods.

In brief, several different general classes of recombination methods are applicable to the present invention and set forth in the references above. First, nucleic acids can be recombined *in vitro* by any of a variety of techniques discussed in the references above, including e.g., DNase digestion of nucleic acids to be recombined followed by ligation and/or PCR reassembly of the nucleic acids. Second, nucleic acids can be recursively recombined *in vivo*, e.g., by allowing recombination to occur between nucleic acids in cells. Third, whole genome recombination methods can be used in which whole genomes of cells or other organisms are recombined, optionally including spiking of the genomic recombination mixtures with desired library components. Fourth, synthetic recombination methods can be used, in which oligonucleotides corresponding to targets of interest are synthesized and reassembled in PCR or ligation reactions which include oligonucleotides which correspond to more than one parental nucleic acid, thereby generating new recombined nucleic acids. Oligonucleotides can be made by standard nucleotide addition methods, or can be made by tri-nucleotide synthetic approaches. Fifth, *in silico* methods of recombination can be effected in which genetic algorithms are used in a computer to recombine sequence strings which correspond to nucleic acid homologues (or even non-homologous sequences). The resulting recombined sequence strings are optionally converted into nucleic acids by synthesis of nucleic acids which correspond to the recombined sequences, e.g., in concert with oligonucleotide synthesis/ gene reassembly techniques. Any of the preceding general recombination formats can be practiced in a reiterative fashion to generate a more diverse set of recombinant nucleic acids. Sixth, methods of accessing natural diversity, e.g., by hybridization of diverse nucleic acids or nucleic acid fragments to single-stranded templates, followed by polymerization and/or ligation to regenerate full-length sequences, optionally followed by degradation of the templates and recovery of the resulting modified nucleic acids can be used.

To illustrate, in one embodiment of the present invention, the shuffling method employed to prepare polynucleotides encoding recombinant derivatizing enzymes comprises:

initiating a polynucleotide amplification process on overlapping segments of a population of variant polynucleotides under conditions whereby one segment serves as a template for extension of another segment, to generate a population of recombinant polynucleotides; and

5                                    selecting or screening a recombinant polynucleotide for a desired property.

The overlapping segments can be prepared by a variety of methods, as described or referenced herein, including, for example, chemical synthesis, cleavage or  
10                                    fragmentation, amplification of the population of polynucleotides, and other methods that are well known in the art.

In another embodiment, the shuffling method used to generate the recombinant derivatizing enzymes comprises:

   hybridizing at least two sets of nucleic acids, wherein a first set of  
15                                    nucleic acids comprises single-stranded nucleic acid templates and a second set of nucleic acids comprises at least one set of nucleic acid fragments; and,

   elongating, ligating, or both, sequence gaps between the hybridized  
nucleic acid fragments, to generate at least substantially full-length chimeric nucleic acid sequences that correspond to the single-stranded nucleic acid templates, thereby recombining  
20                                    the set of nucleic acid fragments, and optionally,

   denaturing the at least substantially full-length chimeric nucleic acid sequences and the single-stranded nucleic acid templates; and

   separating the at least substantially full-length chimeric nucleic acid sequences from the single-stranded nucleic acid templates by at least one separation  
25                                    technique; and, fragmenting the separated at least substantially full-length chimeric nucleic acid sequences by nuclease digestion or physical fragmentation to provide chimeric nucleic acid fragments.

The above references provide these and other basic recombination formats as well as many modifications of these formats. Regardless of the shuffling format which is  
30                                    used, the nucleic acids of the invention can be recombined (with each other or with related

(or even unrelated) to produce a diverse set of recombinant nucleic acids, including, e.g., sets of homologous nucleic acids.

Following recombination, any nucleic acids which are produced can be selected for a desired activity. In the context of the present invention, this can include  
5 testing for and identifying any activity that can be detected in an automatable format, by any of the assays in the art. A variety of related (or even unrelated) properties can be assayed for, using any available assay. These methods are automated according to the present invention as described herein.

DNA mutagenesis and shuffling provide a robust, widely applicable, means  
10 of generating diversity useful for the engineering of proteins, pathways, cells and organisms with improved characteristics. In addition to the basic formats described above, it is sometimes desirable to combine shuffling methodologies with other techniques for generating diversity. In conjunction with (or separately from) shuffling methods, a variety of diversity generation methods can be practiced and the results (i.e., diverse populations of  
15 nucleic acids) screened for in the systems of the invention. Additional diversity can be introduced by mutagenesis methods that are known in the art.

Mutagenesis methods include, for example, those described in Publ. No. WO98/42727; site-directed mutagenesis (Ling et al. (1997) "Approaches to DNA mutagenesis: an overview" In: *Anal Biochem.* 254(2): 157-78; Dale et al. (1996)  
20 "Oligonucleotide-directed random mutagenesis using the phosphorothioate method." *Methods Mol Biol.* 57:369-74; Smith (1985) "In vitro mutagenesis" *Ann. Rev. Genet.* 19, 423-462; Botstein and Shortle (1985) "Strategies and applications of in vitro mutagenesis" *Science* 229, 1193-1201; Carter (1986) "Site-directed mutagenesis" *Biochem J.* 237, 1-7; Kunkel (1987) "The efficiency of oligonucleotide directed mutagenesis" *Nucleic Acids & Molecular Biology*) Eckstein, F. and Lilley, D.M.J. eds Springer Verlag, Berlin)  
25 Mutagenesis using uracil containing templates (Kunkel (1985) "Rapid and efficient site-specific mutagenesis without phenotypic selection" *Proc. Natl. Acad. Sci. USA* 82, 488-492; Kunkel, T.A., Roberts, J.D. & Zakour, R.A. (1987) "Rapid and efficient site-specific mutagenesis without phenotypic selection" *Methods in Enzymol.* 154, 367-382; Bass, S., V. Sorrels, and P. Youderian (1988) "Mutant Trp repressors with new DNA-binding specificities" *Science* 242:240-245); oligonucleotide-directed mutagenesis (for review see,

- Smith, *Ann. Rev. Genet.* 19: 423-462 (1985); Botstein and Shortle, *Science* 229: 1193-1201 (1985); Carter, *Biochem. J.* 237: 1-7 (1986); Kunkel, "The efficiency of oligonucleotide directed mutagenesis" in *Nucleic Acids & Molecular Biology*, Eckstein and Lilley, eds., Springer Verlag, Berlin (1987)); oligonucleotide-directed mutagenesis (*Methods in Enzymol.* 100: 468-500 (1983), and *Methods in Enzymol.* 154: 329-350 (1987); Zoller & Smith (1982) "Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment" *Nucleic Acids Res.* 10, 6487-6500. Zoller & Smith (1983) "Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors" *Methods in Enzymol.* 100, 468-500 Zoller & Smith (1987) "Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template" *Methods in Enzymol.* 154, 329-350) phosphothioate-modified DNA mutagenesis (Taylor et al. (1985) "The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA" *Nucl. Acids Res.* 13: 8749-8764; Taylor et al. (1985) "The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA" *Nucl. Acids Res.* 13: 8765-8787 (1985); Nakamaye and Eckstein (1986) "Inhibition of restriction endonuclease Nci I cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis" *Nucl. Acids Res.* 14: 9679-9698; Sayers et al. (1988), *Nucl. Acids Res.* "Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis" 16:791-802; Sayers et al. (1988) Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide" *Nucl. Acids Res.* 16: 803-814), mutagenesis using uracil-containing templates (Kunkel, *Proc. Nat'l. Acad. Sci. USA* 82: 488-492 (1985) and Kunkel et al., *Methods in Enzymol.* 154:367-382)); mutagenesis using gapped duplex DNA (Kramer et al., "The gapped duplex DNA approach to oligonucleotide-directed mutation construction" *Nucl. Acids Res.* 12: 9441-9456 (1984); Kramer and Fritz, *Methods in Enzymol.* "Oligonucleotide-directed construction of mutations via gapped duplex DNA" 154:350-367 (1987); Kramer et al., *Nucl. Acids Res.* 16: 7207 (1988)); Fritz et al. (1988) "Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro" *Nucl. Acids Res.* 16: 6987-6999 (1988) Mutagenesis using Gapped Duplex DNA; Kramer, W., Ohmayer, A. & Fritz, H.-J. (1988) "Improved enzymatic in vitro reactions in



the gapped duplex DNA approach to oligonucleotide-directed construction of mutations" *Nucleic Acids Res.* 16, 7207; and Bass, S., V. Sorrels, and P. Youderian (1988) "Mutant Trp repressors with new DNA-binding specificities" *Science* 242:240-245).

- Additional suitable methods include point mismatch repair (Kramer et al. (1984) "Point Mismatch Repair" *Cell* 38: 879-887 (1984)), mutagenesis using repair-deficient host strains (Carter et al. (1985) "Improved oligonucleotide site-directed mutagenesis using M13 vectors" *Nucl. Acids Res.* 13: 4431-4443 (1985); Carter (1987) "Improved oligonucleotide-directed mutagenesis using M13 vectors" *Methods in Enzymol.* 154: 382-403), deletion mutagenesis (Eghtedarzadeh and Henikoff (1986) "Use of oligonucleotides to generate large deletions" *Nucl. Acids Res.* 14: 5115), restriction-selection and restriction-selection and restriction-purification (Wells et al. (1986) "Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin" *Phil. Trans. R. Soc. Lond. A* 317: 415-423), mutagenesis by total gene synthesis (Nambiar et al. (1984) "Total synthesis and cloning of a gene coding for the ribonuclease S protein" *Science* 223: 1299-1301; Sakamar and Khorana (1988) "Total synthesis and expression of a gene for the  $\alpha$ -subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)" *Nucl. Acids Res.* 14: 6361-6372; Wells et al. "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites" *Gene* 34:315-323 (1985); and Grundström et al. (1985) "Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis." *Nucl. Acids Res.* 13: 3305-3316), Double-strand break repair (Band aid) (Mandecki (1986) "Oligonucleotide-directed double-strand break repair in plasmids of *Escherichia coli*: a method for site-specific mutagenesis" *Proc. Nat'l. Acad. Sci. USA*, 83:7177-7181). Additional details on many of the above methods can be found in *Methods in Enzymology*, Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis methods.

- Kits for mutagenesis are commercially available. For example, kits are available from, e.g., Stratagene (e.g., QuickChange site-directed mutagenesis kit; Chameleon double-stranded, site-directed mutagenesis kit), Bio/Can Scientific, Bio-Rad (e.g., using the Kunkel method described above), Boehringer Mannheim Corp., Clontech Laboratories, DNA Technologies, Epicentre Technologies (e.g., 5 prime 3 prime kit); Genpak Inc, Lemargo Inc, Life Technologies (Gibco BRL), New England Biolabs, Pharmacia Biotech,

Promega Corp., Quantum Biotechnologies, Amersham International plc (e.g., using the Eckstein method above), and Anglian Biotechnology Ltd (e.g., using the Carter/Winter method above).

In addition, any of the described shuffling techniques can be used in  
5 conjunction with procedures which introduce additional diversity into a genome, e.g. a bacterial genome. For example, techniques have been proposed which produce nucleic acid multimers suitable for transformation into a variety of species, including *E. coli* and *B. subtilis* (see e.g., Schellenberger U.S. Patent No. 5,756,316). When such multimers consist of genes that are divergent with respect to one another, (e.g., derived from natural diversity  
10 or through application of site directed mutagenesis, error prone PCR, passage through mutagenic bacterial strains, and the like), are transformed into a suitable host, an additional source of nucleic acid diversity for DNA shuffling is introduced. Multimers transformed into host species are particularly suitable as substrates for in vivo shuffling protocols.

Alternatively, a multiplicity of polynucleotides sharing regions of partial sequence similarity  
15 can be transformed into a host species and recombined in vivo by the host cell. Subsequent rounds of cell division can be used to generate libraries, members of which, each comprise a single, homogenous population of monomeric or pooled nucleic acid. Alternatively, the monomeric nucleic acid can be recovered by standard techniques and recombined in any of the described shuffling formats.

20 Shuffling formats employing chain termination methods have also been proposed (see e.g., U.S. Patent No. 5,965,408). In this approach, double stranded DNAs corresponding to one or more genes sharing regions of sequence similarity are combined and denatured, in the presence or absence of primers specific for the gene. The single stranded polynucleotides are then annealed and incubated in the presence of a polymerase and a chain  
25 terminating reagent (e.g., uv, gamma or X-ray irradiation; ethidium bromide or other intercalators; DNA binding proteins, such as single strand binding proteins, transcription activating factors, or histones; polycyclic aromatic hydrocarbons; trivalent chromium or a trivalent chromium salt; or abbreviated polymerization mediated by rapid thermocycling; and the like), resulting in the production of partial duplex molecules. The partial duplex  
30 molecules, e.g., containing partially extended chains, are then denatured and reannealed in subsequent rounds of replication or partial replication resulting in polynucleotides which

share varying degrees of sequence similarity and which are chimeric with respect to the starting population of DNA molecules. Optionally, the products or partial pools of the products can be amplified at one or more stages in the process. Polynucleotides produced by a chain termination method, such as described above are suitable substrates for further DNA shuffling according to any of the described formats.

Diversity can be further increased by using non-homology based shuffling methods (which, as set forth in the above publications and applications can be homology or non-homology based, depending on the precise format). For example, incremental truncation for the creation of hybrid enzymes (ITCHY) described in Ostermeier et al. (1999) "A combinatorial approach to hybrid enzymes independent of DNA homology" *Nature Biotechnol.* 17:1205, can be used to generate a shuffled library which can optionally serve as a substrate for one or more rounds of in vitro or in vivo shuffling methods. *See also*, Ostermeier et al. (1999), "Combinatorial protein engineering by incremental truncation," *Proc. Nat'l. Acad. Sci. USA* 96: 3562-3567; Ostermeier et al. (1999), "Incremental truncation as a strategy in the engineering of novel biocatalysts," *Biological and Medicinal Chemistry*, 7: 2139-2144.

Methods for generating multispecies expression libraries have been described (e.g., U.S. Patent Nos. 5,783,431; 5,824,485) and their use to identify protein activities of interest has been proposed (U.S. Patent 5,958,672). Multispecies expression libraries are, in general, libraries comprising cDNA or genomic sequences from a plurality of species or strains, operably linked to appropriate regulatory sequences, in an expression cassette. The cDNA and/or genomic sequences are optionally randomly concatenated to further enhance diversity. The vector can be a shuttle vector suitable for transformation and expression in more than one species of host organism, e.g., bacterial species, eukaryotic cells. In some cases, the library is biased by preselecting sequences which encode a protein of interest, or which hybridize to a nucleic acid of interest. Any such libraries can be provided as substrates for any of the shuffling methods herein described.

In some applications, it is desirable to preselect or prescreen libraries (e.g., an amplified library, a genomic library, a cDNA library, a normalized library, etc.) or other substrate nucleic acids prior to shuffling, or to otherwise bias the substrates towards nucleic acids that encode functional products (shuffling procedures can also, independently have

these effects). For example, in the case of antibody engineering, it is possible to bias the shuffling process toward antibodies with functional antigen binding sites by taking advantage of in vivo recombination events prior to DNA shuffling by any described method.

For example, recombined CDRs derived from B cell cDNA libraries can be amplified and assembled into framework regions (e.g., Jirholt et al. (1998) "Exploiting sequence space: shuffling in vivo formed complementarity determining regions into a master framework" *Gene* 215: 471) prior to DNA shuffling according to any of the methods described herein.

Libraries can be biased towards nucleic acids which encode proteins with desirable enzyme activities. For example, after identifying a clone from a library which exhibits a specified activity, the clone can be mutagenized using any known method for introducing DNA alterations, including, but not restricted to, DNA shuffling. A library comprising the mutagenized homologues is then screened for a desired activity, which can be the same as or different from the initially specified activity. An example of such a procedure is proposed in U.S. Patent No. 5,939,250. Desired activities can be identified by any method known in the art. For example, WO 99/10539 proposes that gene libraries can be screened by combining extracts from the gene library with components obtained from metabolically rich cells and identifying combinations which exhibit the desired activity. It has also been proposed (e.g., WO 98/58085) that clones with desired activities can be identified by inserting bioactive substrates into samples of the library, and detecting bioactive fluorescence corresponding to the product of a desired activity using a fluorescent analyzer, e.g., a flow cytometry device, a CCD, a fluorometer, or a spectrophotometer.

Libraries can also be biased towards nucleic acids which have specified characteristics, e.g., hybridization to a selected nucleic acid probe. For example, application WO 99/10539 proposes that polynucleotides encoding a desired activity (e.g., an enzymatic activity, for example: a lipase, an esterase, a protease, a glycosidase, a glycosyl transferase, a phosphatase, a kinase, an oxygenase, a peroxidase, a hydrolase, a hydratase, a nitrilase, a transaminase, an amidase or an acylase) can be identified from among genomic DNA sequences in the following manner. Single stranded DNA molecules from a population of genomic DNA are hybridized to a ligand-conjugated probe. The genomic DNA can be derived from either a cultivated or uncultivated microorganism, or from an environmental

sample. Alternatively, the genomic DNA can be derived from a multicellular organism, or a tissue derived therefrom.

Second strand synthesis can be conducted directly from the hybridization probe used in the capture, with or without prior release from the capture medium or by a wide variety of other strategies known in the art. Alternatively, the isolated single-stranded genomic DNA population can be fragmented without further cloning and used directly in a shuffling format that employs a single stranded template. Some single-stranded template shuffling formats are described in, for example, WO 98 27239, "METHODS AND COMPOSITIONS FOR POLYPEPTIDE ENGINEERING," Patten et al.; "SINGLE-STRANDED NUCLEIC ACID TEMPLATE-MEDIATED RECOMBINATION AND NUCLEIC ACID FRAGMENT ISOLATION" by Affholter, USSN 60/186,482 filed March 2, 2000; "METHODS FOR GENERATING HIGHLY DIVERSE LIBRARIES," WO 0000632; and "METHOD FOR OBTAINING *IN VITRO* RECOMBINED POLYNUCLEOTIDE SEQUENCE BANKS AND RESULTING SEQUENCES," WO 0009679. In one such method the fragment population derived the genomic library(ies) is annealed with partial, or, often approximately full length ssDNA or RNA corresponding to the opposite strand. Assembly of complex chimeric genes from this population is the mediated by nuclease-base removal of non-hybridizing fragment ends, polymerization to fill gaps between such fragments and subsequent single stranded ligation. The parental strand can be removed by digestion (if RNA or uracil-containing), magnetic separation under denaturing conditions (if labeled in a manner conducive to such separation) and other available separation/purification methods. Alternatively, the parental strand is optionally co-purified with the chimeric strands and removed during subsequent screening and processing steps.

In a conventional approach, single-stranded molecules are converted to double-stranded DNA (dsDNA) and the dsDNA molecules are bound to a solid support by ligand-mediated binding. After separation of unbound DNA, the selected DNA molecules are released from the support and introduced into a suitable host cell to generate a library enriched sequences which hybridize to the probe. A library produced in this manner provides a desirable substrate for further shuffling using any of the shuffling reactions described herein.

It will further be appreciated that any of the above described techniques suitable for enriching a library prior to shuffling can be used to screen the products generated by the methods of DNA shuffling.

In a presently preferred embodiment, the recombinant libraries are prepared  
5 using DNA shuffling. The shuffling and screening or selection can be used to "evolve"  
individual genes, whole plasmids or viruses, multigene clusters, or even whole genomes  
(Stemmer (1995) *Bio/Technology* 13:549-553. Reiterative cycles of recombination and  
screening/selection optionally can be performed to further evolve the nucleic acids of interest.  
Such techniques do not require the extensive analysis and computation required by  
10 conventional methods for polypeptide engineering. Shuffling allows the recombination of  
large numbers of mutations in a minimum number of screening/selection cycles, in contrast  
to traditional, pairwise recombination events. Thus, the sequence recombination techniques  
described herein provide particular advantages in that they provide recombination between  
mutations in any or all of these, thereby providing a very fast way of exploring the manner in  
15 which different combinations of mutations can affect a desired result. In some instances,  
however, structural and/or functional information is available which, although not required  
for sequence recombination, provides opportunities for modification of the technique.

These shuffling methods typically employ at least two variant forms of a  
starting nucleic acid substrate. The variant forms of candidate substrates can show  
20 substantial sequence or secondary structural similarity with each other, but they should also  
differ in at least two positions. The initial diversity between forms can be the result of  
natural variation, e.g., the different variant forms (homologs) are obtained from different  
individuals or strains of an organism (including geographic variants) or constitute related  
sequences from the same organism (e.g., allelic variations). Alternatively, the initial  
25 diversity can be induced, e.g., the second variant form can be generated by error prone  
transcription, such as an error prone PCR or use of a polymerase which lacks proof reading  
activity (*see, e.g.,* Liao (1990) *Gene* 88:107-111), of the first variant form, or, by replication  
of the first form in a mutator strain, or by the mutagenic process of DNase fragmentation and  
reassembly by error prone polymerases. The initial diversity between substrates is greatly  
30 augmented in subsequent steps of recursive sequence recombination.

In a presently preferred embodiment, the shuffling of a "family" of nucleic acids is used to create the library of recombinant polynucleotides. When a family of nucleic acids is shuffled, nucleic acids that encode homologous polypeptides from different strains, species, or gene families or portions thereof, are used as the different forms of the nucleic acids. As genomics provide an increasing amount of sequence information, it is increasingly possible to directly amplify homologs with designed primers. For example, given the sequence of lipase or protease genes from several species, one can design primers for amplification of the homologs. The resulting nucleic acid segments can then be subjected to shuffling.

All of the shuffling methods described herein can be readily employed in the practice of the present invention. For example, in codon modification shuffling (described in detail in "SHUFFLING OF CODON ALTERED GENES" by Patten et al. filed September 29, 1998 (USSN 60/102,362), January 29, 1999 (USSN 60/117,729), and September 28, 1999 (USSN 09/102,362)), nucleic acids are synthesized in which the codons which encode polypeptides are altered, thus making it possible to access a completely different mutational cloud upon subsequent mutation of the nucleic acid. This increases the sequence diversity of the starting nucleic acids for shuffling protocols, which alters the rate and results of forced evolution procedures. Codon modification procedures can be used to modify any derivatizing enzyme encoding nucleic acid herein, *e.g.*, prior to performing DNA shuffling, or codon modification approaches can be used in conjunction with oligonucleotide shuffling procedures as described below.

Codon modification shuffling involves selecting a first nucleic acid sequence that encodes a first polypeptide sequence or portion thereof. A plurality of codon altered nucleic acid sequences, each of which encode part or all of the first polypeptide, or a modified or related polypeptide, is then selected (*e.g.*, a library of codon altered nucleic acids can be selected in a biological assay which recognizes library components or activities), and the plurality of codon-altered nucleic acid sequences is recombined to produce a target codon altered nucleic acid encoding part or all of a second protein. The target codon altered nucleic acid is then screened for a detectable functional or structural property, optionally including comparison to the properties of the first polypeptide and/or related polypeptides. The goal of such screening is to identify a polypeptide that has a structural or functional

property equivalent or superior to the first polypeptide or related polypeptide. A nucleic acid encoding such a polypeptide can be used in essentially any procedure desired, including introducing the target codon altered nucleic acid into a cell, vector, virus (*e.g.*, as a component of a vaccine or immunogenic composition), transgenic organism, or the like.

5            "In silico" shuffling (described in detail in Selifonov and Stemmer in "METHODS FOR MAKING CHARACTER STRINGS, POLYNUCLEOTIDES & POLYPEPTIDES HAVING DESIRED CHARACTERISTICS," filed February 5, 1999 (USSN 60,118,854) and filed October 12, 1999 (USSN 09/416,375)) utilizes computer algorithms to perform "virtual" shuffling using genetic operators in a computer. As applied  
10 to the present invention, derivatizing enzyme gene sequence strings are recombined in a computer system and desirable products are made, *e.g.*, by reassembly PCR of synthetic oligonucleotides. In brief, genetic operators (algorithms which represent given genetic events such as point mutations, recombination of two strands of homologous nucleic acids, etc.) are used to model recombinational or mutational events which can occur in one or more  
15 nucleic acid, *e.g.*, by aligning nucleic acid sequence strings (using standard alignment software, or by manual inspection and alignment) and predicting recombinational outcomes. The predicted recombinational outcomes are used to produce corresponding outcomes. The predicted recombinational outcomes are used to produce corresponding molecules, *e.g.*, by oligonucleotide synthesis and reassembly PCR.

20            In "oligonucleotide-mediated shuffling" (described in Crameri et al. "OLIGONUCLEOTIDE MEDIATED NUCLEIC ACID RECOMBINATION," filed February 5, 1999 (USSN 60/118,813) and filed June 24, 1999 (USSN 60/141,049), and filed September 28, 1999 (USSN 09/408,392)), oligonucleotides corresponding to a family of related homologous nucleic acids (*e.g.*, as applied to the present invention, interspecific or  
25 allelic variants of a derivatizing enzyme encoding nucleic acid) are recombined to produce selectable nucleic acids.

          One advantage of the oligonucleotide-mediated recombination is the ability to recombine homologous nucleic acids with low sequence similarity, or even non-homologous nucleic acids. In these low homology oligonucleotide shuffling methods, one or more set of  
30 nucleic acid segments are recombined, *e.g.*, with a set of crossover family diversity oligonucleotides. Each of these crossover oligonucleotides have a plurality of sequence



diversity domains corresponding to a plurality of sequence diversity domains from homologous or non-homologous nucleic acids with low sequence similarity. The crossover oligonucleotides, which are derived by comparison to one or more homologous or non-homologous nucleic acids, can hybridize to one or more region of the nucleic acid segments, 5 facilitating recombination.

When recombining homologous nucleic acids, sets of overlapping families of oligonucleotides (which are derived by comparison of homologous nucleic acids and synthesis of oligonucleotide segments) are hybridized and elongated (*e.g.*, by reassembly PCR), providing a population of recombined nucleic acids, which can be selected for a 10 desired trait or property. Typically, the sets of overlapping oligonucleotides include a plurality of oligonucleotide member types which have consensus region subsequences derived from a plurality of homologous target nucleic acids. Generally, the sets of overlapping oligonucleotides are provided by aligning homologous nucleic acid sequences to select conserved regions of sequence identity and regions of sequence diversity. A plurality 15 of oligonucleotides are synthesized (serially or in parallel) which correspond to at least one region of sequence diversity.

Sets of segments, or subsets of segments used in oligonucleotide shuffling approaches can be provided by cleaving one or more homologous nucleic acids (*e.g.*, with a Dnase), or, more commonly, by synthesizing a set of oligonucleotides corresponding to a 20 plurality of regions of at least one nucleic acid (typically oligonucleotides corresponding to a full length nucleic acid are provided as members of a set of nucleic acid fragments). In the shuffling procedures described herein, these segments (*e.g.*, segments of derivatizing enzyme encoding nucleic acids) can be used in conjunction with shuffling families of oligonucleotides, *e.g.*, in one or more recombination reaction to produce recombinant 25 derivatizing enzyme encoding nucleic acids.

Often, improvements are achieved after one round of recombination and screening/selection. However, recursive sequence recombination can be employed to achieve still further improvements in a desired property. Sequence recombination can be achieved in many different formats and permutations of formats, which share some common 30 principles. Recursive sequence recombination entails successive cycles of recombination to generate molecular diversity. That is, one creates a family of nucleic acid molecules

showing some sequence identity to each other but differing in the presence of mutations. In any give cycle, recombination can occur *in vivo* or *in vitro*, intracellular or extracellular. Furthermore, diversity resulting from recombination can be augmented in any cycle by applying prior methods of mutagenesis (e.g., error-prone PCR or cassette mutagenesis) to either the substrates or products for recombination. In some instances, a new or improved property or characteristic can be achieved after only a single cycle of *in vivo* or *in vitro* recombination as when using different, variant forms of the sequence, as homologs from different individuals or strains of an organism, or related sequences from the same organism, as allelic variations.

Expression of the recombinant polynucleotides to obtain the recombinant derivatizing enzymes is generally accomplished in cells. The libraries of recombinant polynucleotides can be created either *in vitro* or *in vivo*, as described in US Patent No. 5,837,458. For *in vitro* library generation, the recombinant polynucleotides are thus introduced into cells for expression.

**B. Types of Derivatizing Enzymes Useful for Biocatalytic Synthesis of Combinatorial Libraries**

The methods of the invention are applicable to a wide range of derivatizing enzymes that can catalyze the modification of organic molecules of interest. Such enzymes can modify the substrates by, for example, adding a functional group to the molecule or by modification of an existing functional group on the molecule. Modifications of interest also include addition of chemical moieties onto functional groups. The derivatizing enzymes, in presently preferred embodiments, do not add to the length of the backbone of the organic molecule. Types of reactions of interest are described in, for example, Khmelnsky *et al.* (1996) *Molecular Diversity and Combinatorial Chemistry*, Chapter 14, pp. 144-157 (American Chemical Society), as well as Michels *et al.* (1998) *Tibtech* 16: 210-215. Examples of different types of derivatizing enzymes, and the application of the methods of the invention to these enzymes, are described below.

In addition to the increased diversity of enzymatic activities that are found in the libraries of recombinant enzymes, one can also obtain enzymes that are enhanced in certain properties that increase the usefulness of the enzymes in the modification of organic compounds, such as, natural compounds, non-natural compounds (e.g., 5-fluorouracil,

azidothymidine, etc.), small molecules, and polymers (e.g., peptides and peptide variants, oligonucleotides/polynucleotides and variants thereof, polyhydroxyalkanoates, polysaccharides, polylactic acid, polylactic-co-glycolic acid, polyethylene glycol, and the like). Small molecules employed in the practice of the present invention typically have a  
5 molecular weight of less than about 2500 daltons, usually less than about 2000 daltons, and sometimes less than about 1500 daltons.

These libraries can be screened to identify those library members that encode an enzyme that exhibits an improvement, compared to a wild-type enzyme, in a desired property or properties for use in the reaction of interest. For example, one can screen to  
10 identify those library members that encode an enzyme that has improved substrate specificity for a particular compound, or improved regioselectivity for at a desired functional group on the compound.

In some embodiments, libraries of recombinant derivatizing enzymes are variants of a given wild type gene, into which variation is introduced by diversity generating  
15 methods such as those described herein, e.g., shuffling and gene reassembly shuffling processes. Limited but complete diversity can thus be provided around the given sequence with dense sampling. In other embodiments, the recombination libraries are produced by applying diversity generating methods to several different wild type genes. Limited and incomplete diversity is achieved, which is scattered all over a functional sequence space, as  
20 in sparse sampling. This latter technique is preferred when generating new enzyme specificities.

1. Modification of existing functional groups and introduction of new functional groups into an organic molecule

In some embodiments, the recombinant derivatizing enzymes, and libraries  
25 thereof, can catalyze the modification of an existing functional group that is present on an organic molecule of interest, such as a lead compound. For example, derivatizing agents of interest can oxidize or reduce a functional group, hydrolyze a group, or replace one functional group with another. Other reactions of interest include lactonization, isomerization, and epimerization.

*a. Hydroxylation*

In some embodiments, a hydrogen in an organic molecule is replaced with a hydroxyl group. This can often result in a profound alteration in biological activity.

Hydroxylation is often associated with increased metabolism due to first pass through the liver. Introduction of a hydroxyl group in a drug candidate can also confer a more rapid metabolism by the subsequent action of a group transferring enzyme (*e.g.*, enzymes that catalyze methylation, sulfation, phosphorylation and glycosylation).

Among the derivatizing enzymes that are useful for introduction of hydroxyl groups are the mono- and dioxygenases. A range of monooxygenases known in the art provide appropriate starting points for making libraries of recombinant monooxygenases that are useful in the methods of the invention. One useful class of monooxygenases is exemplified by the heme-dependent eukaryotic and bacterial cytochromes P-450. In the presence of oxygen and an intact redox recycle system, P450s exhibit monooxygenase activity. Addition of hydrogen peroxide or other peroxides, however, can be used to circumvent the NAD(P)H requirement (*i.e.*, allowing for peroxidase activity) toward many of the same substrates. The ability of enzymes such as P450's to perform chemistry at chemically difficult sites is well known. Steroid modification by naturally occurring P450s is widespread in biosynthesis and drug metabolism. Hence, for example, a shuffled library of P450s will generate many new attachment points for further chemical (or enzymatic derivatization) or screening. The other enzyme classes herein mentioned will also have utility in creating new structural diversity in clinically important families of compounds.

The P450 monooxygenase gene family is particularly well suited for use of family shuffling to obtain recombinant derivatizing enzymes. Approximately 70-80 families of P450 monooxygenases are known, from many different species. For identification of homologous genes that can be shuffled together as a family, representative alignments of P450 enzymes can be found in the Appendices of the volume CYTOCHROME P450: STRUCTURE, MECHANISM, AND BIOCHEMISTRY, 2<sup>nd</sup> Addition (ed. by Paul R. Ortiz de Montellano) Plenum Press, New York, 1995) ("Ortiz de Montellano"). An up-to-date list of P450s can be found electronically on the World Wide Web (<http://drnelson.utmem.edu/homepage.html>). To illustrate the application of shuffling to improving a family of P450 enzymes, one or more of the more than 1000 members of this superfamily is selected,

aligned with similar homologous sequences, and shuffled against these homologous sequences. For example, the gene for the bovine P450<sub>sc</sub> enzyme, *CYP11A1*, belongs to a family of closely related P450 genes. DNA shuffling (Cramer *et al.*, *Nature* 391:288) can be used to create hybrid variants from this family of genes, libraries of which can be used to make combinatorial libraries of organic molecule derivatives. Streptomyces, in particular, produces P450 monooxygenases that are used in production of natural products such as antibiotics. Examples of suitable P450 monooxygenase genes for shuffling include the following, each of which is at least 45% identical at the amino acid level:

- cytochrome p450 monooxygenase (*S. venezuelae*) AF087022
- cytochrome p450 monooxygenase (*Sac. erythraea*) M83110
- cytochrome p450 monooxygenase (*Sac. erythraea*) M54983
- cytochrome p450 monooxygenase (*S. hygrosopicus*) X86780
- cytochrome p450 monooxygenase (*S. antibioticus*) L47200

Creation of libraries of recombinant p450 monooxygenase genes is discussed in more detail in co-pending, commonly assigned US Patent Application No. 60/148,850, which was filed on August 12, 1999.

It is noted that the basic chemistry described below with reference to monooxygenases is known. In addition to Ortiz de Montellano, *supra*, a general guide to the various chemistries involved is found in Stryer (1988) BIOCHEMISTRY, third edition (or later editions) Freeman and Co. New York, NY; Pine *et al.* ORGANIC\_CHEMISTRY\_FOURTH EDITION (1980) McGraw-Hill, Inc. (USA) (or later editions); March, ADVANCED\_ORGANIC CHEMISTRY REACTIONS, MECHANISMS and Structure 4th ed J. Wiley and Sons (New York, NY, 1992) (or later editions); Greene, et al., PROTECTIVE GROUPS IN ORGANIC CHEMISTRY, 2nd Ed., John Wiley & Sons, New York, NY, 1991 (or later editions); Lide (ed) (1995) THE CRC HANDBOOK OF CHEMISTRY AND PHYSICS 75TH EDITION (or later editions); and in the references cited in the foregoing. Furthermore, an extensive guide to many chemical and industrial processes applicable to the present invention is found in the KIRK-OTHMER ENCYCLOPEDIA OF CHEMICAL TECHNOLOGY (third edition and fourth edition, through year 1998), Martin Grayson, Executive Editor, Wiley-Interscience, John Wiley and Sons, NY, and in the references cited therein ("Kirk-Othmer").

Other monooxygenase enzymes suitable for introduction of hydroxyl groups and other modifications of organic molecules include those having activities such as alkane oxidation (*e.g.*, hydroxylation, formation of ketones, aldehydes, *etc.*), alkene epoxidation, aromatic hydroxylation, N-dealkylation (*e.g.*, of alkylamines), S-dealkylation (*e.g.*, of reduced thio-organics), O-dealkylation (*e.g.*, of alkyl ethers), oxidation of aryloxy phenols, conversion of aldehydes to acids, alcohols to aldehydes or ketones, dehydrogenation, decarbonylation, oxidative dehalogenation of haloaromatics and halohydrocarbons, Baeyer-Villiger monooxygenation, modification of cyclosporins, hydroxylation of mevastatin, hydroxylation of erythromycin, N-hydroxylation, sulfoxide formation, or oxygenation of sulfonylureas. Other oxidative transformations will be apparent to those of skill in the art. Examples of suitable monooxygenases for use in the invention are described in co-pending, commonly assigned US patent application Ser. No. 09/373,928, entitled "DNA SHUFFLING OF MONOOXYGENASE GENES FOR PRODUCTION OF INDUSTRIAL CHEMICALS," filed August 12, 1999.

Dioxygenases are another class of derivatizing enzymes that are useful for biocatalytic synthesis of organic molecule derivatives. The bacterial arene dioxygenases (ADOs), for example, can oxidize  $\pi$ -bonds to the corresponding vicinal diols. In the presence of oxygen, and of a reducing compound such as NAD(P)H, these enzymes catalyze the reductive dioxygenation of compounds as diverse as aromatic rings and non-aromatic multiple bonds. The non-phenolic nature of ring *cis*-dihydroxylation products arising from action of arene dioxygenases offers a significant advantage for manufacturing organic molecule derivatives by avoiding the accumulation toxic and reactive epoxide intermediates which may significantly impair the performance of the biocatalyst.

Arene dioxygenases include, for example, toluene 2,3-dioxygenase, isopropylbenzene 2,3-dioxygenase, benzene-1,2-dioxygenase, biphenyl-2,3-dioxygenase, naphthalene-1,2-dioxygenase, and many homologous and/or functionally similar enzymes. Suitable arene dioxygenase-encoding polynucleotides can be obtained from many organisms using cloning methods known to one skilled in the art. The following list provides examples of polynucleotides that encode arene dioxygenases and are suitable for use in the methods of the invention. The loci are identified by GenBank ID and encode complete or partial protein components of the arene dioxygenases. Suitable loci include, for example: [PSETODC1C]

toluene-1,2-dioxygenase; [AF006691], [PJU53507], [PSECUMA], [REU24277]  
 isopropylbenzene-2,3-[E04215], [PSEBDO] dioxygenase; benzene-1,2-dioxygenase;  
 [AEBPHA1F], [CTU47637], [D78322], [D88020], [D88021], [PSEBPHA], [PSEBPHABC],  
 [PSEBPHABCC], [PSU95054], [RERBPHA1], [RGBPHA], [RSU27591] biphenyl-2,3-  
 5 dioxygenase; [PSU15298] chlorobenzene dioxygenase; [AB004059], [AF010471],  
 [AF036940], [AF053735], [AF053736], pAF079317], AF004283], [AF004284],  
 [PSENAPDOXA], [PSENAPDOXB], [PSENDOABC], [PSEORF1], [PSU49496]  
 naphthalene-1,2-dioxygenase; [AF009224], [PSEBEDC12A] benzoate-1,2-dioxygenase;  
 [PWWXYL] toluate dioxygenase; [ASCBAABC], [U18133] 3-chlorobenzoate-3,4-  
 10 dioxygenase; [PCCBDABC] 2-chlorobenzoate-1,2-dioxygenase; [BSU62430] 2,4-  
 dinitrotoluene dioxygenase; [PSU49504] 2-nitrotoluene dioxygenase; [PPU24215] p-  
 cumate-2,3-dioxygenase; [PSEPHT] phthalate-4,5-dioxygenase; [AB008831], [ACCANI],  
 [D85415] aniline 1,2-dioxygenase; [D90884] phenylpropionic acid 2,3-dioxygenase;  
 [PPPOBAB] phenoxybenzoate dioxygenase; [AF060489], [AB001723], and [D89064]  
 15 carbazole dioxygenase.

Also of utility are organisms whose genomes contain genes encoding other  
 dioxygenases, including tetralin-5,6-dioxygenase, Sikkema *et al.*, *Appl. Environ. Microbiol.*  
 59:567-573, (1993); *p*-cumate-2,3-dioxygenase DeFrank *et al.*, *J. Bacteriol.* 129:1356-1364  
 (1977); fluorenone 1,1a-dioxygenase, Selifonov *et al.*, *Biochem. Biophys. Res. Comm.*  
 20 193:67-76(1993); dibenzofuran-4,4a dioxygenase, Trenz *et al.*, *J. Bacteriol.* 176:789-795  
 (1994); phthalate-3,4-dioxygenase, Eaton *et al.*, *J. Bacteriol.* 151:48-58 (1982); and 2-  
 chlorobenzoate-1,2-dioxygenase (Selifonov *et al.*, *Biochem Biophys. Res. Comm.*  
 213(3):759-767 (1995), and the like. These and other dioxygenases that are suitable for use  
 in making the enzyme libraries of the invention are described in co-pending, commonly  
 25 assigned US patent application Ser. No.60/148,450, entitled "DNA SHUFFLING OF  
 DIOXYGENASE GENES FOR PRODUCTION OF INDUSTRIAL CHEMICALS," which  
 was filed August 12, 1999.

Once a hydroxyl group has been introduced into a lead compound or other  
 organic molecule, it is often desirable to add a functional group to the hydroxyl (*e.g.*,  
 30 glycosylation, acylation, and the like), as described below. Accordingly, the invention also  
 provides methods in which a library of organic molecule derivatives obtained by contacting

the organic molecule with a first library of recombinant derivatizing enzymes is subsequently contacted with a second library of recombinant derivatizing enzymes. The enzymes of the second library are often, but not necessarily, those that catalyze the addition of a chemical moiety to a functional group. Alternatively, the hydroxylated compound can be modified by chemical or other means that are known to those of skill in the art.

*b. Halogenases*

Halogenases constitute another example of a class of derivatizing enzyme that can be used to obtain libraries of organic molecule derivatives. The halogenases generally halogenate aromatic rings that can become part of complex natural or non-natural products and other organic molecules that are of interest as, for example, lead compounds. Examples of suitable halogenases include the following: halogenase PmA, PmB, PmC (U74493; *P. fluorescens*), putative halogenase P1tM, P1tD, P1tA (AF081920; *P. fluorescens*), putative oxygenase/halogenase (Y16952; *Amycolatopsis orientalis*). Although these particular enzymes have less than about 35% amino acid sequence identity, the polynucleotides that encode the enzymes are useful as probes to obtain more closely related halogenases that can be used for DNA shuffling.

*c. Other substitutions*

Similarly, one can introduce a sulfur-containing group into an organic compound. Thiols, for example, are generally introduced in order to generate a thiolate anion, which have a strong affinity for heavy metals. Often, heavy metals are found in enzyme active sites. Derivatizing enzymes that are useful for these embodiments include, for example, the aryl sulfotransferase family. This family of enzymes can be used to transfer a sulfo group onto the aromatic part of an organic molecule. The aryl sulfotransferase family includes many members that have very high amino acid sequence identity (>80%), such that they can be readily shuffled together to generate the libraries of recombinant derivatizing enzymes. Examples of suitable sulfotransferase genes that can be used for recombination include, for example, arylamine sulfotransferase (U33886; *Homo sapiens*), phenol sulfotransferase (D85541; *Macaca fascicularis*), phenol sulfotransferase (D29807; *Canis familiaris*), phenol sulfotransferase (U34753; *Bos taurus*), and minoxidil sulfotransferase (L19998; *Rattus norvegicus*).



In additional embodiments, one or more basic groups are substituted for preexisting functional groups. The basic groups most typically used in medicinal chemistry are the amines, the amidines, the guanidines, and almost all nitrogen-containing heterocycles. Introduction of such groups into a molecule that already has biological activity  
5 has essentially the same solubilizing effect as introduction of an acid function. Amines and basic heterocycles are virtually ubiquitous in successful drugs. One can readily introduce an amine by, for example, use of an acyltransferase or esterase using a bifunctional compound that includes an amine.

## 2. Addition of chemical moieties onto functional groups

10 Additional embodiments of the invention provide recombinant derivatizing enzymes, and libraries thereof, that can catalyze the addition of one or more chemical moieties onto functional groups that are present on an organic molecule of interest, such as a lead compound. In these embodiments, the recombinant derivatizing enzymes of the invention are those that can attach a group to the core functional drug moiety at a position  
15 that does not destroy function of the drug. Such attachments can increase the solubility of the drug moiety, as a prodrug, for example.

The attachment can be either reversible or irreversible. Reversible attachments include, for example, attachment of esters, peptides, and glucosides. Irreversible attachments include, for example, attachments via O- and N- alkylation. Creation of C-C  
20 bonds can be achieved using grafted side chains (*e.g.*, dimethylaminoethyl or morpholinoethyl chains) or acidic side chains (*e.g.*, carboxylic, sulfonic,  $-\text{OSO}_3\text{H}$ ,  $-\text{PO}_3\text{H}_2$ ,  $-\text{OPO}_3\text{H}_2$ ), or with neutral groups (*e.g.*, glyceryl). Larger solubilizing groups can also be added using the enzymes and methods of the invention. Examples of these include, but are not limited to,  $-\text{O}-\text{CH}_2-\text{CH}_2-\text{COOH}$ ,  $-\text{NH}_2-\text{CH}_2-\text{CH}_2-\text{CH}_2-$ ,  $-\text{C}=\text{N}-\text{O}-\text{CH}_2-\text{CO}_2\text{H}$ , O-  
25 morpholinoethyl- and  $-\text{O}-\text{CO}-\text{CH}_2-\text{CH}_2-\text{CO}_2\text{H}$ .

Nonionizable side chains, including, for example, hydroxylated and polyoxymethylene side chains or diverse glucosides, can also be attached in order to enhance solubility. This class of side chains also includes polyethylene glycol derivatives, which are also used for increased solubility as well as sustained release.

30 Examples of derivatizing enzymes that are useful for addition of a chemical moiety to a preexisting functional group on a lead compound or other organic molecule are

glycosyltransferases, acyltransferases, amidases, N-methyltransferases, phosphotransferases, aryl sulfotransferases, and the like.

*a. Acyltransferases*

Acylation is one type of modification chemistry that could theoretically  
5 provide much diversity in derivatization of organic molecules. Traditional chemical processes for acylation, however, are typically non-selective and require multiple protection and de-protection steps. Enzymatic acylation in organic solvent by acyltransferases, including lipases and proteases, for example, can provide certain advantages such as substrate-, stereo- and regio-selectivity. However, it is unlikely that one could obtain, from a  
10 set of naturally occurring acyltransferases one that will possess the desired variety of substrate-, stereo-, or regio-specificity for any particular organic molecule. Therefore, the present invention provides libraries that contain a multitude of recombinant acyltransferases that can be used to synthesize acylated derivatives of lead compounds and other organic molecules.

15 Thus, the invention provides libraries of recombinant polynucleotides that encode lipase and protease enzymes, and acyltransferases. These methods involve the creation of libraries of recombinant polynucleotides using as substrates polynucleotides that encode enzymes that can carry out an acylation reaction. Such enzymes include, for example, lipases and proteases. The reverse reaction of lipases and proteases in organic  
20 solvent can transfer various acyl groups onto hydroxyl sites of the complex natural products. Those enzymes usually possess broad substrate specificity but low activity.

Families of lipases, for example, can readily be identified from publicly available databases. One example of an lipase family that is suitable for shuffling (amino acid identity greater than 50%) includes the following members: Y00557, *Vibrio cholerae*;  
25 D50587, *Pseudomonas sp* KFCC10818 (AAD22078), *Pseudomonas aeruginosa* (BAA23128), *P. aeruginosa* (D50587); *Acinetobacter calcoaceticus* (AF047691); and *P. wisconsinensis* (U88907 and 2072017), *Pseudomonas sp* (P26877), *Bacillus subtilis* (M74101); *Bacillus pumilus* (A34992); *Galactomyces geotrichium* (A02813); *Candida rugosa* (WO 99/14338); and *Acinetobacter calcoaceticus* (S61927).

30 Many genes that encode acyltransferases which use various carboxylic acid derivatives of coenzyme A as substrates are known, and enzymes catalyzing these reactions

are ubiquitous in prokaryotic and eukaryotic organisms. Examples of nucleic acids that are suitable for use as substrates include, for example, galactoside 6-O acetyl transferase (EC 2.3.1.18); *lacA* of *E. coli* (B0342 (*lacA*) or of other organisms (GENBANK loci MG396;D02\_orf152 (*lacA*); MJ1064 (*lacA*), MJ1678, MTH1067); serine O-

5 acetyltransferase (EC 2.3.1.30, (GENBANK locus B3607 (*cysE*), HI0606 (*cysE*), HP1210 (*cysE*), SLR1348 (*cysE*)); alcohol O-acetyltransferase (EC 2.3.1.84), from, for example, *Saccharomyces cerevisiae* (loci YGR177C, YOR377W); arylamine N-acetyltransferase (EC 2.3.1.118, representative GENBANK loci include Q00267, D90786, Z92774, I78931, AF030398, AF008204, AF042740); carnitine O-acetyltransferase (EC 2.3.1.7), from, for

10 example, mammalian or yeast origin (GENBANK loci YAR035(YAT1), and YM8054.01(CAT2)); choline O-acetyltransferase (EC 2.3.1.6), *e.g.*, that of mammalian origin; and acetyl CoA:deacetylindoline 4-O-acetyltransferase (EC 2.3.1.107) (St-Pierre *et al.* (1998) *Plant J.* 14: 703-713).

Suitable acyl donors for the improved enzymes of the invention include, for

15 example, those compounds that can serve as a donor for the particular enzymes. Representative acyl donor substrates include vinyl esters, trifluoroethyl esters and other aliphatic esters, as well as benzyl and fatty acids, and the like. *See, e.g.*, Mozhaev *et al.* (1998) *Tetrahedron* 54: 3791-3982, in particular p. 3976.

In a preferred mode of this invention, acyl transferase genes that are shuffled

20 are those that encode enzymes which provide transfer of the acetyl group, and use endogenous pool of acyl-CoA compounds in the cell of the host microbial strain. The endogenous pool of acyl-CoA can also be enhanced by introduction of an acyl-CoA ligase, optionally improved by DNA shuffling, into host microbial strain that carries out the acylation reaction. The strain is then supplied with exogenous acetate or other carboxylic

25 acid in the medium, which is then attached to CoA by the acyl ligase. Suitable acyl ligases and methods for their optimization are described in co-pending, commonly assigned US patent application Ser. No. 09/373,928, entitled "DNA SHUFFLING OF MONOOXYGENASE GENES FOR PRODUCTION OF INDUSTRIAL CHEMICALS," filed August 12, 1999.

30 Compounds of interest for derivatization by acylation include, for example, natural products and such as polyketides, flavonoids, peptide antibiotics, and the like, as well

as non-naturally occurring compounds. Such compounds find use as, for example, antibiotics, chemotherapeutic agents, and the like. Generally, the substrate molecules have one or more hydroxyl residues at which acylation can occur. Regioselectivity is particularly important for molecules that have multiple functional groups at which acylation can occur.

- 5 The methods of the invention provide a means by which one can obtain an enzyme that acylates the functional group or groups of interest, but not other groups that might otherwise be susceptible to acylation.

Acylation of specific molecules can alleviate unfavorable properties.

- Anticancer drugs, including those that act by disrupting microtubulin dynamics, are among  
10 the compounds for which the methods of the invention are useful for developing derivatives of the drugs that have improved properties. These compounds include, for example, colchicine, colcemid, podophylloxotoxin, taxol, vinblastine, vincristine, and the like. One particular example of a substrate of interest is epothilone, which is a potent anticancer drug candidate that is currently in the research stage. Selective acylation of two hydroxyl groups  
15 on this compounds can increase its water solubility. The recombinant acyltransferase libraries of the invention can be used to obtain derivatives that are specifically acylated at these positions. Additional examples are rapamycin and FK506. Acylation of the C-28 hydroxyl group of rapamycin or the undehydrated C-35 hydroxyl of FK506 can be used to separate their immunosuppressive activities from their nerve regenerative activities (Gold,  
20 B.G. (1997) *Mol. Neurobiol.* 15: 285-306). It is known that the part of rapamycin or FK506 binding to FKBP (FK binding protein) is responsible for the neuroregenerative activity. Acylation can destroy the binding of the FKBP-Rapamycin (or FK506) to the effector protein (calcineurin). Therefore, acylation of the aforementioned hydroxyl groups will disrupt the calcineurin binding. Regio selectivity will play a major role in these  
25 modifications, since there are several hydroxyl groups in both molecules.

- The screening of the libraries of recombinant polynucleotides that encode lipases, proteases, or other acylating enzymes, whether obtained by DNA shuffling or other methods as described above, is done most easily *in vitro* using purified or partially purified enzymes or bacterial or yeast lysates in organic solvent systems, by one or more of the  
30 screening methods described below. For example, one can detect increased formation of acylated derivatives of natural products and small molecules by detecting physical

differences between the substrates and the derivatives arising from the enzyme-catalyzed reactions. These methods include HPLC, mass-spectrometry, UV/Vis and IR spectroscopy, NMR, and the like.

Another presently preferred method uses a labeled acyl-donor precursor, *e.g.* labeled carboxylic acid or its derivative, administered to the cells that express libraries of genes that encode shuffled lipases, proteases, or other acyltransferases. The amount of label in the reaction products is measured. For hydrophobic reaction products, one can extract the derivatives into a suitable organic solvent, or one can use solid-phase extraction of these compounds by addition of a sufficient amount of hydrophobic porous resin beads (*e.g.*, XAD 1180, XAD-2, -4, -8). In the case of a radiolabel, scintillating dye can be present in the organic solvent, added to the samples, or chemically incorporated in the bead polymer. The latter constitutes a modification of scintillation proximity assay method.

The methods for detection regioselectivity of the acylation reactions include, for example, HPLC, and in an HTP modality, flow-through NMR spectroscopy. When NMR spectroscopy is used for determination of relative amounts of different regiomer acylated derivatives of the natural products or small molecules, the later are preferably obtained by action of the enzymes on isotopically ( $^{13}\text{C}$  and/or  $^2\text{H}$ ) labeled substrates. Another variation of the NMR technique includes use of isotopically labeled precursors of acyl donor intermediates.

#### *b. Glycosyltransferases*

Another example of a derivatizing enzyme of interest for generating combinatorial libraries of organic molecule derivatives are the glycosyltransferases. Glycosylation can increase bioavailability, reduce toxicity and increase water solubility of organic molecules, including lead compounds. Because glycosylations are difficult to perform chemically, novel sugar containing antibiotics, such as new glycopeptide and glycosylated macrolide antibiotics, are difficult to make.

Using glycosyltransferases, however, allows one to accomplish glycosylation of organic acceptor compounds that contain one or more hydroxyl groups. Therefore, with the greater variety in glycosylation ability provided by the recombinant enzyme libraries of the invention, many variants of organic molecules are obtainable. With the technology provided herein, new enzymes are provided that can catalyze a variety of previously

unavailable glycosylations. For example, the recombinant derivatizing enzymes in the libraries of the invention can exhibit changed specificity for both acceptors (*e.g.*, complex natural and synthetic organic molecules) and donors (*e.g.*, different sugars). Increased ability to synthesize aminodeoxy sugars can also be obtained, *e.g.*, by biotransformation. Using the recombinant derivatizing enzymes of the invention, new substrates can be accessed, new enzymatic activity can be created and improved; difficult chemical processes can be replaced by biocatalysis, and high scale ups can be accomplished.

Glycosyltransferases can be evolved using the diversity generating methods described herein, including, for example, shuffling, to generate recombinant glycosyltransferases that exhibit optimal performance with respect to a variety of different reaction parameters. Typical reaction parameters include, but are not limited to, specificity of reaction, degree of promiscuity of enzymes and stereochemistry. For example, the enzymes are optionally evolved to transfer different nucleotide diphosphate (NDP) sugars and NDP-sugar analogs; to transfer sugars to different acceptor molecules; to attach sugars at different positions compared to naturally occurring enzymes, to possess ambiguity towards positions in multiple site containing acceptors, and to catalyze multiple step-wise glycosylations.

In another embodiment, enzymes can be evolved to generate recombinant derivatizing enzymes that utilize alternative sugars which are optionally synthetic. For example, activated sugars, such as desoxy and sulfated sugars; non-natural sugars, *e.g.*, nitrosylated, sulfonated, phosphonated, and dideoxy sugars; polyalcohols, *e.g.*, inositol, inositol-phosphates, and inositol phosphonates; other sugar like structures and compounds and alternative nucleotides.

Recombinant glycosyltransferases are also optionally used to transfer sugars to alternative sugar receptors, including but not limited to polyketides, non-ribosomal peptides, complex molecules from organic synthesis, and libraries of chemical compounds. Other sugars acceptors of interest in the present invention include, but are not limited to, aglycosyl vancomycin hydrochloride (a peptide antibiotic), somatostatin (a growth hormone), insulin and glucagon-release inhibitor, cholic acid (a detergent steroid), nogalamycin (an anti-tumor antibiotic), L-thyroxine (a thyroid hormone), syringaldazine, aclarubicin (an anti-tumor antibiotic and commercial RNA synthesis inhibitor), ritodrine HCl

(an adenergic agonist and smooth muscle relaxant), rifamycin (an antibiotic), and ristomycin sulphate (an antibiotic). Each of these compounds has 3-dimensional similarity to vancomycin aglycone, as defined by the molecular dynamics interface with the Available Chemical Database that is available through Chemweb (<http://www.chemweb.com/> databases). These compounds and their sugar attachment points of interest are shown in Figures 1-10. Other natural products of interest for glycosylation include, for example, lovastatin, aglycosyl erythromycin, echinocandin, taxol and cephalixin.

Any molecule which contains at least one hydroxyl group is optionally glycosylated with an evolved glycosyltransferase. Pharmacologically interesting compounds are preferred. Sugar acceptors with more than one hydroxy group are optionally glycosylated at only one of the positions. Thus different isomers can be produced by glycosylating at one or the other of the positions. Alternatively, compounds with more than one hydroxy group are optionally glycosylated at different positions to a different extent, when NDP sugars are limiting for example. In yet another embodiment, compounds are treated multi dimensionally with combinations of NDP-sugars and glycosyltransferases, providing iterative glycosylation.

In some embodiments, the glycosyltransferases are selected from those which transfer hexose residues from UDP-hexose derivatives. Preferred hexoses include, for example, D-glucose, D-galactose and D-N-acetylglucosamine. Sugars of interest in attachment using evolved glycosyltransferases include, but are not limited to, the following: UDP-N-acetylgalactosamine, UDP-N-acetylglucosamine, UDP-galactose, UDP-galacturonic acid, UDP-glucuronic acid, UDP-mannose, UDP-xylose, UDP-glucose, TDP-glucose, CDP-glucose, ADP-glucose, ADP-ribose, ADP-mannose, GDP-fucose, GDP-glucose, and GDP-mannose, all of which are available from Sigma (St, Louis, MO). Deoxy sugars, such as 2-deoxy-D-xylo-hexose, 2-deoxy-D-arabino-hexose, L-fucose, L-rhamnose, D-mycinoside, L-vallarose, D-fucose, D-quinovose, D-rhamnose, D-canarose, D-oliiose, D-digitose, D-boivinose, L-oleandrose, chalcose, D-amicetose, L-rhodinose, ascarylose, abequose, paratose, tyvelose, colitose, and the like. These sugars and others are described in *Annu. Rev. Microbiol.* 48, 223-256 (1994).

The invention provides methods of obtaining recombinant polynucleotides that encode glycosyltransferase enzymes that are enhanced in certain properties that increase

more of several known methods. The following are illustrative examples of glycosyltransferase-encoding nucleic acids that can be used as source nucleic acids for creation of the recombinant libraries which are then screened to identify those that exhibit an improvement in the glycosylation of organic compounds, such as altered substrate  
5 specificity. For example, inositol 1- $\alpha$ -galactosyltransferase, EC 2.4.1.123; phenol beta-glucosyltransferase, EC 2.4.1.35 (NTU32643, NTU32644); flavone 7-O-beta-glucosyltransferase, EC 2.4.1.81; flavonol 3-O-glucosyltransferase, EC 2.4.1.91 (AB002818, ZMMCCBZ1, AF000372, AF028237, AF078079, D85186, ZMMC2BZ1, VVUFGT); o-dihydroxycoumarin 7-O-glucosyltransferase, EC 2.4.1.104; vitexin beta-glucosyltransferase,  
10 EC 2.4.1.105; coniferyl-alcohol glucosyltransferase, EC 2.4.1.111; monoterpenol beta-glucosyltransferase, EC 2.4.1.127; arylamine glucosyltransferase, EC 2.4.1.71; sn-glycerol-3-phosphate 1-galactosyltransferase, EC 2.4.1.96; glucuronosyltransferase, EC 2.4.1.17 (RNUDPGTR, AA912188, AA932333); the human UGT and isoenzymes (~35 genes); salicyl-alcohol glucosyltransferase, EC 2.4.1.172; 4-hydroxybenzoate 4-O-beta-D-  
15 glucosyltransferase, EC 2.4.1.194; zeatin O-beta-D-glucosyltransferase, EC 2.4.1.203; D-fructose-2-glucosyltransferase, VFAUDPGFTA; and ecdysteroid UDP-glucosyltransferase (egt) MBU41999 may all be used as substrates for creation of the recombinant libraries of the invention.

Additional suitable glycosyltransferase genes can be found in many  
20 microorganisms which one skilled in the art can isolate from various soil, sediment, air and aqueous samples by enrichment culture techniques. Glycosyltransferases specifically isolated from the soil bacteria glycosylate several of polyketide aglycones and such glycosylated natural products possess many different biological activities, such as antibiotic, and anticancer. Genes coding for such enzymes are readily available from the public  
25 database. For example, glycosyltransferases (*S. antibioticus*, AJ002638; *Sac erythraea*, Y14332; *S. venezuelae*, AF079762; *S. peucetius*, L47164 and *S. fradiae*, X81885). Those genes share more than 50% of the amino acid sequence identity and any two or more are thus ideal for shuffling together as a family.

As an example, glycosyltransferases that are used for initial shuffling are  
30 *gtfA*, *gtfB*, *gtfC*, *gtfD*, and *gtfE*, from different *Amycolatopsis orientalis* strains. These genes code for glycosyltransferases that transfer sugar moieties to the aglycons of vancomycin and



the usefulness of the enzymes in the synthesis of glycosylated organic compounds. In presently preferred embodiments, polynucleotides that encode the improved glycosyltransferase enzymes are introduced into microorganisms that are added to the biocatalytic reaction mixture. In some embodiments, the glycosyltransferase is expressed by  
5 a microorganism species other than that from which the glycosyltransferase gene was obtained.

In presently preferred embodiments, the glycosyltransferases used in the methods of the invention are optimized by subjecting nucleic acids that encode the enzymes to recombination and subsequent selection to identify those recombinant polynucleotides  
10 that encode enzymes having an enhanced property of interest. For example, one can select for those recombinant polynucleotides that encode enzymes that can selectively glycosylate at only one hydroxyl group, that can control regioselectivity to provide either of two possible isomeric compounds, or that are capable of glycosylating a wider variety of compounds, such as enzymes that utilize a variety of sugars and sugar analogs not normally utilized by  
15 naturally occurring glycosyltransferases, and enzymes that glycosylate a variety of organic compounds to which naturally occurring glycosyltransferases are unable to attach a sugar molecule.

Libraries of recombinant polynucleotides that are subjected to selection or screening to identify those that encode recombinant glycosyltransferases having enhanced  
20 properties can be created by application of, for example, the various recombination-based diversity generating methods described herein (such as shuffling), to nucleic acids that encode these enzymes (i.e., the nucleic acids are the substrates for recombination). Sources of glycosyltransferase genes that are suitable for use as substrates in the creation of the libraries of recombinant polynucleotides include, for example, the *gtf* genes from *A.*  
25 *orientalis* that encode glycosyltransferases that catalyze, e.g., the transfer of glucose to aglycosyl vancomycin. Enzymes that catalyze these reactions are ubiquitous in prokaryotic and eukaryotic organisms.

One or more glycosyltransferases can be selected from the glycosyltransferase superfamily, aligned with similar homologous sequences, and shuffled  
30 against these homologous sequences. Glycosyl transfer reactions are ubiquitous in nature, and one of skill in the art can isolate such genes from a variety of organisms, using one or

eremomycin, which are non ribosomal peptide antibiotics. Zmijewski & Briggs *FEMS Microbiology Letters* 59, 129-134 (1989). Solenberg *et al. Chem & Biol.* 4, 195-202 (1997). Wageningen *et al. Chem. & Biol.* 5, 155-162 (1998). For example, GtfB and gtfE transfer glucose from TDP-glucose or UDP-glucose onto vancomycin. The glycosyltransferase genes share similarities between 59% (*gtfA-gtfD*) and 82% (*gtfB-gtfE*). The protein sequences share similarities between 52% (*gtfA-gtfD*) and 80% (*gtfB-gtfE*). The five published genes can be amplified from different *Amycolatopsis orientalis ssp orientalis* strains (*gtfD* and *gtfE* from ATCC 43490 or ATCC 43491 and *gtfA*, *gtfB*, *gtfC* from NNRL 18098). Another number of uncharacterized but related glycosyltransferases genes are optionally PCR amplified from other *A. orientalis* strains, *e.g.* ATCC 19795, 21425, 35164, 15165, 15166, 39444, 43333, 53550, and 53630, and cloned into a suitable cloning and expression vector. Further genes can be amplified from the balhimycin producer *Amycolatopsis mediterranei* DSM5908 (Pelzer *et al. (1997) J. Biotechnol.* 57: 115-128), and from other *Amycolatopsis* strains. The expression of *gtf*-encoded proteins in *E. coli* can be tested by either SDS-PAGE and Coomassie stain and/or if a detection tag was added by Western blot. Single clones, *e.g.*, of *gtfB* and *gtfE*, can be tested for their wild type activity. For example, *gtfB* and *gtfE* transfer glucose from TDP-glucose or UDP-glucose onto the aglycon of vancomycin. Folena-Wassermann *et al. J. of Antibiotics* 39, 1395-1406 (1986). The *in vitro* glucosylation of the vancomycin aglycon can be monitored by reverse phase HPLC. Solenberg *et al. Chem. & Biol.* 4, 195-202 (1997). Subsequently, functional *gtfB* and *gtfE* clones and several clones of other genes, *e.g.*, *gtfA*, *gtfC*, *gtfD* and the like, expressing a polypeptide chain of the desired size are used to generate PCR products of the *gtf* genes in the context of a screening vector. DNaseI fragments of each PCR product are generated and reassembled, *e.g.*, by a variety of shuffling methods as described above. Typically the fragment size is between 25 base pairs and 250 base pairs, but this size is easily determined experimentally by methods well known in the arts.

### *c. Methyltransferases*

The methyltransferases are another example of a derivatizing enzyme of interest that can add a chemical moiety onto a functional group present on a lead compound or other organic molecule. S-adenosylmethionine (SAM) dependent methyltransferases (MTs), for example, make up a class of enzymes which form methyl-ester, methyl-ether,

methyl-thioether, methyl-amine, and methyl-amide derivatives of proteins, nucleic acids, sugars, polysaccharides, lipids, lignin, and a variety of low molecular weight compounds (such as macrolides). SAM carries an activated methyl group that is efficiently transferred to nucleophiles having a broad range of chemical reactivity. Transfer of the activated methyl  
5 group from SAM to the recipient nucleophile is thermodynamically favorable, thereby driving the methyl transfer reaction essentially to completion.

One class of methyltransferases of interest are the N-methyltransferases. As an example, the following N-methyltransferases have at least 59% amino acid sequence identity, thus making the family particularly well suited for shuffling: putative TDP-N-  
10 dimethyldesosamine-N-methyltransferase (U77459; *Saccharomyces erythraea*), methyltransferase (AJ002638; *S. antibioticus*), N,N-dimethyltransferase (AF079762; *S. venezuelae*), N-methyltransferase (X81885; *S. fradiae*). This family of enzymes usually methylates the amine group of the amino deoxy sugars attached to complex natural products.

Also of interest are the O-methyltransferases, several families of which are  
15 known. For example, the following family of methyltransferases can methylate the hydroxyl groups of complex natural products: 31-demethyl-FK506 methyltransferase (U65940; *Streptomyces sp*), methyltransferase (X86780; *Streptomyces hygroscopicus*), carbomycin 4-O methyltransferase (D30759; *Streptomyces thermotolerans*), and O-methyltransferase (M93958; *Streptomyces mycarofaciens*). These family members are greater than 45%  
20 identical at the amino acid level.

#### d. Amidases

The invention also provides recombinant libraries of amidases. This family of enzymes may be used to introduce amide groups into organic molecules. The reverse of the amidase reaction converts carboxylic acid groups into a carboxylic acid amide. One such  
25 family that is suitable for use in the methods of the invention includes the following amidases, which are at least 55% identical at the amino acid level: N-acetyl-anhydromuramyl-L-alanine amidase (AF082575; *Pseudomonas aeruginosa*), N-acetyl-anhydromuramyl-L-alanine amidase (U40785; *Enterobacter cloacae*), AmpD protein (X15237; *E. coli*) and AmpD protein (U32716; *Haemophilus influenzae* Rd).

*e. Phosphotransferases*

The addition of a phospho group onto an existing functional group of a lead compound or other organic molecule is also of interest. Thus, the invention provides libraries of recombinant phosphotransferases that are useful for obtaining phosphorylated organic molecule derivatives. As an example, the macrolide and peptide phosphotransferase family, members of which have at least 36% amino acid sequence identity, can be subjected to recombination (*e.g.*, macrolide 2'-phosphotransferase I (D16251; *E. coli*), macrolide 2'-phosphotransferase II (D85892; *E. coli*), viomycin phosphotransferase (X02393; *S. vinaceus*)). This group of enzymes transfer a phospho group onto either macrolide or peptide antibiotics as way to inactivate them. Through using libraries of recombinant phosphotransferases, one can obtain phosphorylation of different sites of the macrolides or peptide antibiotics.

*f. Other enzyme classes*

Enzyme classes other than the ones listed above are also very important in terms of introducing or modifying functional groups in lead generation or/and optimization. For example, enzymes capable to catalyze oxidation-reduction reactions are important to oxidize functional alcohols to aldehydes/ketones or reduce aldehydes/ketones groups to alcohols in organic compounds. These newly created groups can then be further modified by other classes enzymes as described. One such family suitable for shuffling is that of lactate dehydrogenase, which converts ketone to alcohol with >80% amino acid sequence identity: (Y00711, *Homo sapiens*; U07181, *Rattus norvegicus*; 77022A, *Sus scrofa domestica*; L79954, *Trachemys script*, etc.). Alcohol dehydrogenase is another family enzyme which oxidize alcohol group into aldehyde. Suitable genes with this enzyme family are readily available for shuffling (M84409, *Homo sapiens*; L15704, *Peromyscus maniculatus*; 156882, *Struthio camelus*; P80222, *Alligator mississippiensis*, etc). Shuffling of these two families of enzymes can change their substrate specificity towards more complex organic compounds.

Other enzyme families such as enzymes capable to oxidize sulfides to sulfoxides, thiols to thioaldehydes and enzymes capable to catalyze cyanohydrin formations and epoxidations etc are also targets for DNA shuffling, therefore a valuable catalysts for use in combinatorial biosynthesis.

**C. *Use of Recombinant Derivatizing Enzyme Libraries to obtain Combinatorial Libraries of Organic Molecule Derivatives***

The invention provides, in additional embodiments, methods for obtaining a library of organic molecule derivatives. These methods involve contacting an organic molecule (a substrate) with a library of recombinant derivatizing enzymes and other necessary reactants to form the library of organic molecule derivatives. The derivatizing enzymes, as described above, catalyze a reaction such as: a) modification of one or more functional groups present on the organic molecule; b) addition of a chemical moiety onto one or more functional groups present on the organic molecule; or c) introduction of a new functional group onto the organic molecule.

**1. Organic molecules of interest for derivatization**

Organic molecules of interest include, for example, those that have pharmacological activity, herbicide or pesticide activity, and the like. Among the organic molecules of interest are natural products, such as antibiotics (including, for example, polyketides, steroids, non-ribosomal peptide antibiotics, and the like). Steroids for example, are an extremely widely used basic structure for drugs whereby the substituents on the rings target the drug to many different therapeutic targets. Most of these are derived from natural sources and screened for efficacy. Substituents observed on steroid drugs include hydroxyls, methoxy, alkoxy, glycosylations, sulfations, halogenations, double and triple bonds, carbonyls, and the like. The chemical derivatization of the steroid ring structure is readily achieved at a few well described sites or by modification of the naturally occurring structures or non-naturally occurring variants thereof.

Cyclic glycopeptides and macrolides such as vancomycin and erythromycin are also chemically difficult structures that can be modified by the application of shuffled enzyme libraries. There are many such structures isolated from nature and described in the literature, and in company vaults, that have interesting bioactivities but fail in other regards. toxicity, bioavailability, solubility, pharmacokinetics, lack of selectivity are some of the reasons drug candidates are unable to become drugs. Application of the shuffled libraries can be used to improve these and other characteristics.

Prostaglandins, alkaloids, anthraquinones are other families of molecules which have many biologically active members. These are also good candidates for improvement with shuffled enzyme libraries.

Specific examples of pharmaceutical compounds that one can derivatize using the recombinant derivatizing enzymes include, for example, tubocurarine chloride, alcuronium chloride, pancuronium bromide, vecuronium bromide, atracurium besilate, 776C85, 7C1Me-MDO-CPT, 9-aminocamptothecin, A-007, A-108835, A-121798, purpurea glycosides A and B, lanatosides A, B and C,  $\alpha$ -acetyldigoxin,  $\beta$ -acetyldigoxin, digoxin,  $\beta$ -methyl digoxin, k-strophanthoside, k-strophanthin- $\beta$ , convallaside, convallatoxin, glucoscillaren A, scillaren A, proscillaridin, scillarenin. Also of interest are choleric and cholekinetic drugs, including, for example, hymecromone, febupol, chenodeoxycholic acid and ursodeoxycholic acid. Fluocortolone, paramethasone, dexamethasone, betamethasone, cortisone, hydrocortisone, prednisone, prednisolone, triamcinolone acetonide, triamcinolone, methylprednisolone and prednylidene are among the glucocorticoids that are suitable for derivatization. Corticosteroids of interest also include, for example, prednicarbate, hydrocortisone aceponate, fluocortinbutyl, ioteprednol etabonate, and the like.

## 2. Enzymatic Reactions

To obtain the libraries of organic molecule derivatives, the substrates are contacted with the members of the library of recombinant enzymes. The enzymatic reactions can be performed in numerous ways, including the use of whole cell biotransformation, permeabilized cells, cell lysate, and purified protein, for example.

Whole cell biotransformation occurs when the substrate (*e.g.*, an organic molecule) is exposed to cells containing the library of recombinant derivatizing enzymes. The library can be expressed as a surface protein on a replicable genetic package, *e.g.*, phage or yeast display, or as a secreted protein that interacts with the substrate in solution. The enzymes can also be expressed inside the cell, in which case the substrate will diffuse into the cell before the reaction occurs. In each case, the resulting product of the derivatizing enzyme activity is isolated from the cells by methods known to those of skill in the art, including, for example, centrifugation, precipitation, extraction with organic solvents, and filtration.

The cells that express the library can be permeabilized by addition of a number of well known permeabilizing agents such as polymyxin B sulfate. The level of permeabilizing agent can be modified to allow the passage of substrate and product to freely diffuse to the enzymes of the library and out of the cell again. At higher levels of  
5 permeabilizing agent the protein may be released into solution. The compounds of interest will be isolated as for whole cells.

The library can be used as a cell lysate, whereby the cells expressing the library are broken by addition of well known lysis conditions which includes addition of detergent, PMBS and lysozyme, or sonication. The cell debris may be removed before  
10 reaction by centrifugation though this may not be necessary. Substrate is then added to the lysate and after an incubation at a defined temperature and for a defined length of time. The product is then extracted as before and analyzed as described below.

Alternatively, the recombinant derivatizing enzymes encoded by the library can be purified by many well known techniques before screening or use to make derivatives  
15 of organic molecules. Such methods include, for example, gel filtration, ion exchange, affinity, or hydrophobic chromatography to yield either partially or fully purified protein. Many other purification methods are known to those of skill in the art. The purified protein is then exposed to the substrate under conditions that favor enzyme activity.

The reaction conditions used for the transformation are optimized for  
20 maximal enzymatic turnover by standard methods, which include the use of optimal salt levels, buffer, temperature, and length of reaction. The substrate, and any other substrates consumed in the enzymatic reaction, are preferably used at a concentration that promotes a high turnover rate.

The contacting of an organic molecule and other reactants with a recombinant  
25 derivatizing enzyme can be done using the entire library of enzymes at once, or with pools of recombinant enzymes from the library, or with a single recombinant enzyme in each reaction. If a pool is used, the pool can be deconvoluted to isolate the particular clone that exhibits a desired activity once an active pool had been identified using the described methods. For example, colonies that express each member of the library of recombinant  
30 derivatizing enzymes can be placed in microtiter plates or other suitable container and subjected to high throughput screening.

In some embodiments, the members of the library of recombinant enzymes are immobilized on a solid support prior to contacting with the other reactants. For example, the recombinant polynucleotides that encode the enzymes can be introduced into an expression vector that also includes a coding sequence for a tag, such that the recombinant derivatizing enzymes are expressed as a fusion protein with a tag. Alternatively, a tag can be attached to the derivatizing enzymes after their expression. The tag is typically a member of a binding pair for which a corresponding member is readily obtainable and immobilizable on a solid support. For example, the recombinant enzyme can be expressed as a fusion with biotin, which can then be immobilized by binding to streptavidin. Other suitable binding pairs include, for example, maltose binding protein and amylose, histidine tags and an immobilized metal ion, glutathione-S-transferase and reduced glutathione, streptavidin binding tags and streptavidin, epitope tags (*e.g.*, E-tag, myc-tag, HAG-tag, His-tag) and corresponding antibodies, chitin binding domains and chitin, S-tag and RNase minus S-peptide mutant, cellulose binding proteins and domains and cellulose, thioredoxin and DsbA and a thiol compound (*e.g.*, Thiobond™), poly-cationic tags (*e.g.*, poly-arginine) and a poly-anion column, IgG and IgG-derived peptides and protein A, protein G, and the like, calmodulin binding peptide and calmodulin, histactophilin and immobilized metal chelate chromatography.

The member of the binding pair to which the tag attached to the enzymes binds is preferably attached to a solid support. Solid supports suitable for use are known to those of skill in the art. As used herein, a solid support is a matrix of material in a substantially fixed arrangement. Exemplar solid supports include glasses, plastics, polymers, metals, metalloids, ceramics, organics, etc. Solid supports can be flat or planar, or can have substantially different conformations. For example, the substrate can exist as particles, beads, strands, precipitates, gels, sheets, tubing, spheres, containers, capillaries, pads, slices, films, plates, dipsticks, slides, *etc.* Magnetic beads or particles, such as magnetic latex beads and iron oxide particles, are examples of solid substrates that can be used in the methods of the invention. Magnetic particles are described in, for example, US Patent No. 4,672,040, and are commercially available from, for example, PerSeptive Biosystems, Inc. (Framingham MA), Ciba Corning (Medfield MA), Bangs Laboratories (Carmel IN), and



BioQuest, Inc. (Atkinson NH). The substrate is chosen to maximize signal to noise ratios, primarily to minimize background binding, for ease of washing and cost.

Separation of the recombinant enzymes from other cellular components, or from reactants and the like, can be effected for example, by removing a bead or dipstick from a reservoir, emptying or diluting a reservoir such as a microtiter plate well, rinsing a  
5 bead (*e.g.* beads with iron cores may be readily isolated and washed using magnets), particle, chromatographic column or filter with a wash solution or solvent. The separation step will sometimes include an extended rinse or wash or a plurality of rinses or washes. For example, where the solid substrate is a microtiter plate, the wells may be washed several times with a  
10 washing solution, which typically includes those components of the reaction mixture that can interfere with subsequent screening of the organic molecule derivatives, such as salts, buffer, detergent, nonspecific protein, etc.

The libraries of recombinant derivatizing enzymes provided by the invention are useful not only to obtain libraries of organic molecule derivatives, but also provide a  
15 source from which one can identify a recombinant enzyme that catalyzes a particular reaction of interest. For example, once a particular organic molecule derivative is identified as having a desired property, one can identify a particular recombinant enzyme from the enzyme library that can catalyze the formation of the particular derivative.

### 3. Screening of organic molecule derivatives

20 The libraries of recombinant derivatizing enzymes are useful for the production of combinatorial libraries of organic molecule derivatives, which are in turn screened to identify those that exhibit a desired activity. In these embodiments, the product of the screening is often a compound that had not previously been made. In addition, the libraries of recombinant enzymes provide a source from which one can identify an enzyme  
25 that catalyzes a particular known modification of an organic molecule. Thus, for example, one can obtain from the library an enzyme that makes possible enzymatic synthesis of a known compound that previously could only be synthesized by less efficient methods, such as chemical synthesis.

Once a library of organic molecule derivatives has been synthesized, the  
30 library is generally subjected to screening to identify those derivatives that are of particular interest. Generally, to identify a derivative that exhibits an improvement in a particular

biological activity, one can use a bioassay that is designed to allow detection and/or quantitation of the desired activity. Screening for desired biological activity (including, for example, cell toxicity, genotoxicity, and the like), desired bioavailability (including properties such as plasma half-life, renal clearance, and the like), desired physicochemical property (including properties such as, water solubility, lipid solubility, solubility in organic solvent (e.g., n-octanol), water solubility, pH stability (e.g., the low pH environment of the stomach), temperature stability, resistance to intestinal enzymes, resistance to hepatic enzymes, resistance to plasma enzymes, tissue permeability (e.g., dermal, mucosal, and the like), blood-brain barrier permeability), and other desired properties that can be achieved by derivitization, can all be conducted randomly, *e.g.*, without regard to the structures of the compounds, or can be preceded by analysis of the structures of the compounds in the library to identify those that have a particular structure of interest. Once compounds having the desired biological activity have been identified, structural analysis can be employed to identify the structural features imparted by the library of recombinant derivatizing enzymes.

In most cases, the recombinant derivatizing enzymes present in the library are expected to chemically modify a given substrate in a predictable fashion. For example, a glycosyltransferase will transfer a sugar moiety onto an amine or hydroxyl of the substrate. This will lead to predictable changes in the physical behavior of the molecule, which can be utilized for screening. The chemical transformation catalyzed by a particular library on any given substrate is liable to be the same, *e.g.*, glycosyltransferases will place a sugar onto the substrate, methyltransferases will add a methyl group, P450's will tend to add a hydroxyl group, *etc.* This allows generic screening methods to be devised for each library. For example, a glycosyl transferase will always produce a sugar-substrate linkage and so a specific chemical test for a linked sugar will detect product formation. A kinase library would transfer a phosphate group onto the substrate and specific phosphate tests will detect the presence of product.

A number of analytical screening tools are available for determining the structure of compounds in a combinatorial library. For example, a number of methods are known that are capable of detecting low concentrations of compounds in a high throughput format, including flow analysis NMR and mass spectrometry. These analytical tools, or others including UV/Vis and IR spectroscopy, fluorescence spectroscopy, luminescence, and

the like, can be used to both detect and quantify the novel compounds produced in the enzymatic reactions.

One hundred percent turnover of the substrate to product is not expected in a library screen and so the analytical techniques are preferably set up to detect the specific changes produced by the enzymatic activity. For example, the presence in a library of recombinant enzymes of an enzyme that has methyltransferase activity on a particular substrate of interest could be detected by observation of an increase of 14 amu in the mass spectrum after contact with the enzyme. Thus, the changes in the chemical structure of the substrate caused by the library can often be specifically monitored and detected. These can then be correlated to the member of the library of recombinant enzymes that catalyzed the particular reaction.

Another approach to detect the presence in an enzyme library of a recombinant enzyme that catalyzes a particular reaction upon a new substrate, for example, is the incorporation of a molecular marker during the course of reaction. Suitable labels include, for example, radiolabels such as  $^3\text{H}$ ,  $^{14}\text{C}$ ,  $^{32}\text{P}$ , and the like. This can be achieved using radioactive co-substrates such as  $^3\text{H}_3$  methyl S-adenosyl methionine, whereby only the methylated product of reaction will be labeled. Other labels can also be used; many are known in the art. For example, glycosylation can be detected by use of a sugar molecule that includes a label.

In certain instances the product of the action of the shuffled library upon the substrate is expected to provide a product that is more stable than the substrate towards external stress such as extremes of pH, or increase the solubility of the compound in a particular solvent. This change in behavior can also be monitored by suitable analytical or bioassay methods.

In some cases, the detection of the newly formed product may require separation of the product from the substrate by standard chromatographic methods such as TLC, HPLC, CE, or GC. This can be followed by spectroscopic or other (*e.g.*, flame ionization, mass spectrometry) methods to detect the formation of a novel compound of interest.

## EXAMPLES

The following examples are offered to illustrate, but not to limit the present invention.

### Example 1

#### 5     **Generation of glycosyltransferase enzyme libraries and high throughput screening for the production of desvancosamine vancomycin**

This Example describes how one can generate a library of recombinant glycosyltransferases and use the enzymes for the production of desvancosamine vancomycin.

#### 10    **A.     Cloning of the gtfA, B, C, D, E genes from *Amycolatopsis orientalis ssp orientalis* strains**

##### ***1.     Generation of the gtf encoding DNA***

##### **Preparation of genomic DNA**

*Amycolatopsis orientalis ssp. orientalis* strains ATCC43490 and NRRL  
15    18098 are obtained from ATCC and NRRL. Initial cultures on agarose petri dishes are prepared according to the supplier's recommendation. Liquid cultures are grown for two to five days in TSB at 25°C -28°C. The genomic DNA is extracted according to a standard procedure (Ausubel *et al.* (1987) *Current Protocols in Molecular Biology*, 1<sup>st</sup> Edn., John Wiley & Sons, Inc., NY).

##### 20             **PCR by add-on-primer**

PCR is performed using genomic DNA, 1 pmol of gene specific primer, 200 μM dNTPs, 2 units Deep Vent Polymerase and 0.2 units of its 5'-3' exonuclease activity lacking variant in the presence of 1.5 M betaine and 1-3.5 mM MgSO<sub>4</sub> in a 50 μl volume according to the enzyme supplier's (New England Biolabs) instructions. In all cases hot start  
25    using wax beads (MBP) is employed. On a DNA engine thermal cycler, the cycles are set to the following scheme: 95°C for 5 min initially; 5 cycles: 95°C 45 sec, 76°C 1min 20sec; 5 cycles: 95°C 45sec, 75°C 1min 20sec; 5 cycles: 95°C 45sec, 74°C 1min 20sec; 10 cycles: 95°C 45sec, 73°C 1min 20sec; 10 cycles: 95°C 45sec, 73°C 1min 20sec. All primers are designed according to the sequence entry U84349 and U84350. For the amplification of  
30    gtfA, the primers gtfA.For and gtfA.Rev are used. For the amplification of gtfB, the primers

gtfB.For and gtfB.Rev are used. For the amplification of gtfC, the primers gtfC.For and gtfC.Rev are used. For the amplification of gtfD, the primers gtfD.For and gtfD.Rev are used. For the amplification of gtfE, the primers gtfE.For and gtfE.Rev are used (Table 1).

**Table 1****Primers, Oligonucleotides, Polynucleotides**

Primer	SEQ ID NO.	Sequence
GtfA.For	1	AGGAGATATACATATGCGCGTGTTGATTACGGGGTGTGGA TCGCGC
GtfB.For	2	AGGAGATATACATATGCGTGCTGTTGGCGACGTGTGGA TCGCGC
GtfC.For	3	AGGAGATATACATATGCGTGTTGTTGTGCGACGGCTGGC AGCCGC
GtfD.For	4	AGGAGATATACATATGCGTGTTGTTGTGCGGTGTGCGGA ACCCGC
GtfE.For	5	AGGAGATATACATATGCGTGTTGTTGTGCGACCTGTGGG AGCCG
GtfA.Rev	6	ACCACCACCTTCGATATCGGAACCGGCGGGAACAGTCGGC TTTTC
GtfB.Rev	7	ACCACCACCTTCGATATCGGAACCGGCGGAAACAGTCGGC TTTTC
GtfC.Rev	8	ACCACCACCTTCGATATCGGAACCGGCGGAGAACAGCCGAC TTTTC
GtfD.Rev	9	ACCACCACCTTCGATATCGGAACCGGCGGGAACGGCGGG CTCGTT
GtfE.Rev	10	ACCACCACCTTCGATATCGGAACCGGCGGGAACGGCGGG CTGGTT
Bio-Seq	11	CCTCTCCTTTGCTAGCCATCAGATTTCCCCTCGTGCTTTC
CKFor1	12	CGAATTTCTAGAGAAGGAGATATACATATG
CKFor2	13	CCCCAGGCTTTACACTTTATGCTTCCGGCT
CKFor3	14	GGTACCCGATAAAAGCGGCTTCCTGACAGG
BirFor	15	TGGCATGGATGAGCTCTACAAATAAGGAGACAATTCATG AAGGAT
BirRev	16	GGGCTTATTTTTCTGCACTACGCA
H3Rev	17	AATTTAAGGGTAAGTTTCCGTATGTAG
N1Rev	18	CATGCCGTTTCATCTGATCCGGATAAC

The resulting PCR products are digested with *NdeI* and *EcoRV*. The digested PCR product that corresponds to the gtf gene is purified by agarose gel electrophoresis and QIAEXII (Qiagen).

## 2. *Properties and structure of the vector pCKZEBB.*

The vector pCKZEBB is derived from pAK400 (Krebber *et al.* (1997) *J. Immunol. Meth.* 201: 35-55. The following features of pAK400 are kept. The lacI<sup>q</sup> gene is kept for repression of the lac operon, the transcriptional terminator (hp<sup>1</sup>) between lacI<sup>q</sup> gene and lac promoter (lac<sup>p/o</sup>) is kept to terminate read through transcription from the lacI promoter into the lac promoter controlled operon reducing basal non-induced expression, the lac promoter operator was kept for transcription initiation and transcription control, the T7g10 leader from T7 phage gene10 in front of the target gene start codon was kept to enable strong translation initiation from the ATG start codon in the NdeI restriction site. Behind the NdeI-HindIII lac promoter operator controlled expression cassette there is the lpp transcriptional terminator (lppt) encoded followed by the fl origin of replication to allow single stranded DNA production followed by the chloramphenicol resistance gene (cam<sup>R</sup>), and the ColE1 origin for double stranded DNA replication.

In pCKZEBB a lac promoter operator controlled polycistronic message replaces the lac promoter operator controlled monocistronic message in pAK400. The lac promoter transcribed operon is located between the unique NdeI and HindIII of the pAK400 vector. In pCKZEBB a variant of the lacZ gene (start codon ATG incorporated in NdeI site, internal NdeI removed, EcoRV site added to end of gene in front of stop codon, resulting EcoRV lacZ piece inverted in vector) is inserted as a stuffer fragment in the NdeI EcoRV target gene cloning site. This lacZ fragment will be replaced by the target glycosyltransferase genes. Behind lacZ there is a biotinylation tag encoded (aa sequence) followed by the translational coupling tag derived from the end of the trpB gene. Both tags are fused in frame to the target glycosyltransferase gene when it replaces the lacZ stuffer fragment. The A nucleotide of the stop codon of the translational coupling tag (TGA) constitutes part of the translational start codon of a green fluorescent protein-encoding gene (GFP; Crameri *et al.* (1996) *Nature Biotechnol.* 14: 315-319)). The GFP gene is followed by the birA gene PCR cloned including a ribosomal binding site from BL21(DE3). There seems to be a sequence ambiguity in the birA gene as there exists an NcoI restriction site in this region. A map of pCKZEBB is shown in Figure 19, and the nucleotide sequence of the vector is shown as

SEQ ID NO:19.

*E. coli* transformed with pCKZEBB do not turn green fluorescent when grown on 30 µg/ml chloramphenicol and 1 mM IPTG. When the stuffer lacZ fragment is replaced by the full-length target gene in frame with the biotinylation-translational coupling tag, A), the IPTG induced expression of the target gene turns the plasmid harboring bacteria green fluorescent by translational coupling to the GFP gene (Oppenheim & Yanofsky (1980) *Genetics* 95, 785-795) and, B), the target gene will be biotinylated *in vivo* by the biotinylation tag (Schatz (1993) *Bio/Technology* 11: 1138-1143) via the birA derived biotin holoenzyme ligase (Smith *et al* (1998) *Nucl. Acids Res.* 26: 1414-1420).

### 3. *Cloning of the gtF PCR's into pCKZEBB.*

The vector pCKZEBB is cut with NdeI and EcoRV removing the lacZ gene stuffer as two parts. The resulting vector is dephosphorylated by using calf intestinal phosphatase. The DNA fragment corresponding to the vector is isolated from agarose gels by QIAEXII (Qiagen). The above mentioned EcoRV and NdeI digested PCR product is ligated into the vector fragment according to standard procedures. After ligation *E. coli* TG1 electrocompetent cells (Stratagene) are electroporated with the ligation and plated on LB-Agar plates containing 30 µg/ml chloramphenicol and 1 mM IPTG and grown overnight at 37°C. Green fluorescent colonies showing different extents of fluorescence are picked and plasmid DNA is prepared.

### 4. *Restriction analysis and sequencing*

The resulting vectors are analyzed by restriction enzymes and clones that contain inserts are sequenced. Plasmids that express one of the glycosyltransferases as a biotinylation-translational coupling tag fusion protein are identified. Clones harboring genes that correspond to the published sequence are used as template for shuffling.

## B. *Recombination and mutation of single, double, triple, quadruple and all five genes combinations by family shuffling*

### 1. *Amplification of wt genes in pCK vector.*

The glycosyltransferase genes are amplified from the resulting plasmids, including some vector derived flanking regions by primers CK.For3 and H3.Rev using a

polymerase according to the manufacturer's recommendations. The PCR is purified by Qiaquick columns (Qiagen).

**2. Generation of random DNA fragments.**

PCR product derived from either the plasmids are digested with DNaseI (Boehringer). The reaction is stopped on dry ice and the fragments in the desired size range are isolated from 2% agarose gels using glassfilter disks (Whatman) and dialysis membranes (Spectrapor) (Stemmer (1994) *Proc. Nat'l. Acad. Sci. USA* 91: 10747-10751 and Stemmer (1994) *Nature* 370: 389-391).

**3. Assembly of glycosyltransferase genes.**

For each family assembly reaction several concentrations and ratios of DNaseI DNA fragments and PCR cycling parameters are adjusted so that in step 4 a maximal amount of shuffled genes are obtained (Crameri *et al.* (1998) *Nature* 391: 288-291, Christians *et al.* (1999) *Nature Biotechnol.* 17: 259-64).

**4. Rescue of glycosyltransferase genes by PCR**

Two µl of the final assembly reaction is used as template. In the final PCR reaction, 1 µM primer CK.For2 and N3.Rev, 0.2 mM each nucleotide of 1 unit of Tag polymerase are added. The following PCR parameters are set: 1 cycle, 96°C 3 min; 30 cycles, 96°C 0.5 m, 60°C 0.5 m, 72°C 1.5 m; 1 cycle, 72°C 5 min.

**C. Cloning of the gtf PCR products into pCKZEBB**

The expression vector pCKZEBB and the PCR rescued shuffled glycosyltransferase genes are digested with XbaI and EcoRV. The vector pCKZEBB is in addition dephosphorylated. The vector fragment and the glycosyltransferase encoding PCR fragment are isolated from agarose gels and are ligated with each other.

Electrocompetent *E. coli* TG1 is transformed with the ligation mix and after 1 hour shaking at 37°C plated on LB-agar containing 30 µg/ml Chloramphenicol, 1% Glucose and grown overnight at 37°C.



**D. Prescreening, generation of master plates, and expression of the glycosyltransferase library**

Colonies are picked into LB-Cam-Glucose and grown ON at 37°C to generate the master plate. From the master plates colonies are arrayed onto LB-Cam-IPTG-Agar and the plates are incubated overnight at 37°C. Green fluorescent colonies are identified by exposure of the plate to 365 nm ultraviolet light. The respective green fluorescent colonies from the master plate are re-arrayed into 96 well plates each well filled with 100 µl 2YT-Cam 30-1%Glucose and grown overnight at 37°C. 50 µl culture are transferred to 1 ml of 2YTCam30- 1mg/ml biotin and grown for 7 h at 16°C. Then 50 µl of 100 µM IPTG is added and the cultures are grown overnight at 16°C.

**E. Lysis of the cells by a combination of lysozyme and Polymyxin B sulfate**

The cultures are centrifuged (4000 rpm for 15 minutes) to pellet the cells. The cell pellets are washed with 500 µl of 50 mM ammonium formate (pH 7.4) and pelleted once more. The cells are resuspended in 300 µl lysis buffer (10 µL Ready to Lyse lysozyme (Epicentre), 2 µL RNase A (Qiagen), 2 µL DNase I (Boehringer), 2 µL 1M MgSO<sub>4</sub>, in 10 ml of 1 mg/ml Polymyxin B sulfate (Sigma), 2 mM DTT in 50 mM ammonium formate pH 7.4) and agitated at ambient temperature for thirty minutes. The lysate is then clarified by centrifugation (15 minutes at 4000 rpm).

**F. Purification of the proteins from single clones by magnetic beads.**

Streptavidin coated magnetic beads are arrayed into 96 well plates. The beads are washed, using the beads' magnetic properties, with buffer (50 µM ammonium formate pH 7.4, 2 mM DTT) and resuspended in 20 µl of buffer per well. Clarified cell lysate (100 µL) is transferred to the beads from the lysis plate and incubated for 15 minutes at ambient temperature. The beads are then washed five times with buffer (150 µL) and finally resuspended in 20 µl buffer.

**G. Performing *in vitro* modification of compounds by glycosyltransferases from the library**

Reaction mixture (80 µL) is added to the purified proteins on the beads and the beads are agitated at ambient temperature overnight. Reaction mixture contains, 150µM vancomycin aglycone (synthesized as described in *J. Chem. Soc. Chem. Commun.* (1988)

1306-1307), 500  $\mu$ M UDP glucose, 2 mM DTT in 50 mM ammonium formate pH 7.4. The reactions are quenched by addition of 1 volume of methanol and the mixture is centrifuged (5 minutes at 2000 rpm). Supernatant (100 $\mu$ L) is withdrawn to a new 96 well plate and subjected to mass spectrometry.

5    **H.    Measuring the occurrence of glycosylation**

The quenched reaction mixture (10  $\mu$ l) is injected into a triple quadrupole electrospray mass spectrometer set in the positive mode. Molecular ions are allowed to pass through the first quadrupole (1143 amu for vancomycin aglycone, 1305 amu for desvancosamine vancomycin) and subjected to collision in the second quadrupole before  
10 peak detection of the daughter ions at 100 amu in the third quadrupole. Integration of the peaks obtained from this process are directly proportional to product formation. This determines the relative fitness of the library clones in the production of desvancosamine vancomycin.

**I.    Recursive use of the procedure**

15            If desired, these steps can be repeated. For example, one can repeat steps B to H using multiple genes that encode variants of a particular derivatizing enzyme, using single genes obtained from a library, using single genes shuffled with wild-type genes for backcrossing, and with multiple genes, each of which encodes an enzyme having a different activity.

20            In a variation of the procedure the UDP-glucose in step G is replaced by other NDP-sugars. The MS parameters in step H are adapted to detect the predicted molecular ions.

**Example 2**

**Generation of a methyltransferase library and evolution of an erythromycin 6-O-  
25            methyltransferase for production of clarithromycin**

This Example describes the generation of a library of recombinant O-methyltransferases (OMTase) and the use of enzymes from the library to synthesize derivatives of clarithromycin (6-O-methyl erythromycin).

A family of erythromycin analogs having a 6-methoxy group have been shown to have useful pharmaceutical properties. These compounds are presently prepared by a multi-step chemical methylation of erythromycin A and its analogs (Figure 11). An enzyme capable of selectively transferring an activated methyl group to the 6-hydroxyl group would allow for a one step high yield production of this class of erythromycin analogs *in vivo* or as a single bioconversion *in vitro*. This Example describes an approach for obtaining such methyltransferases.

No erythromycin 6-OMTase activity has been detected at this time. Thus it is necessary to create an OMTase of novel specificity. The chances of finding a new activity by sampling  $10^4$  -  $10^5$  members of a shuffled library are greatly increased if the sequence diversity of the library originates from naturally occurring sequences rather than from random point mutations. Such a library spans a larger portion of sequence space and is enriched with functional sequences. Therefore, DNA shuffling is performed using a family of homologous genes encoding OMTases that specifically methylate substrates similar to the 6-hydroxyl of erythromycin. Since it is uncertain which members of this family will be more influential in the generation of 6-OMTase activity, a variety of shuffled libraries are generated. For example, each of the subfamilies is shuffled alone, as well as shuffling the entire family together. This is accomplished using several shuffling formats that are designed to effect the recombination of genes of both high and low sequence identity.

S-adenosylmethionine (SAM) dependent methyltransferases (MTs) make up a class of enzymes that form methyl-ester, methyl-ether, methyl-thioether, methyl-amine, and methyl-amide derivatives of proteins, nucleic acids, sugars, polysaccharides, lipids, lignin, and a variety of low molecular weight compounds (such as macrolides). SAM carries an activated methyl group that is efficiently transferred to nucleophiles having a broad range of chemical reactivity. Transfer of the activated methyl group from SAM to the recipient nucleophile is thermodynamically favorable thereby driving the methyltransfer reaction essentially to completion (Figure 12). A family of seven genes is known that encode SAM-dependent OMTases specific for secondary alcohols on carbomycin, midecamycin, saframycin, rapamycin, rifamycin, and FK506 (Figure 13). A comparison of these substrate nucleophiles with the 6-hydroxyl of erythromycin A suggests that only minor adjustments in local specificity would be required for the parent OMTases to accept erythromycin as

substrate. Another gene of interest is that which encodes ERYG, which O-methylates the mycarose moiety of erythromycin C, resulting in the synthesis of erythromycin A. EryG shares 54% identity at the DNA level with rapQ perhaps providing an additional subfamily of OMTases containing tertiary alcohol OMTase activity (Figure 14).

5                   The genes to be shuffled are synthesized either from genomic DNA or from synthetic oligonucleotides by the PCR. These genes are then cloned into a suitable vector for expression. The complete sequence of the gene encoding the carbomycin-4-OMTase is not known, but one can clone the gene or the partial sequence can be shuffled with the full sequences of the other OMTases.

10                  Several libraries of SAM dependent OMTases are generated. These libraries are screened against erythromycin A and its analogs for 6-OMTase activity. The identified clones are pooled and evolved further to improve the enzyme to a practical level of activity.

                  Generally,  $10^4$ - $10^5$  clones from the family shuffled library are screened to identify those that have deserythromycin 6-OMTase activity. Cell cultures are grown in the  
15                  presence of deserythromycin A, and the supernatants of these cultures are then removed and assayed for the presence of 6-O-methyl deserythromycin A oxime. The OMTase genes from the identified clones are isolated, pooled, shuffled, and then screened for increased deserythromycin A 6-OMTase activity. Additional cycles of shuffling and screening will continue until the enzyme activity has reached a level suitable for production of 6-O-methyl  
20                  deserythromycin.

                  To insure the identification of useful activities, the shuffled library can be screened for 6-OMTase activity against erythronolide B, deserythromycin A, erythromycin A, and their oxime derivatives. While it is possible that no deserythromycin A oxime 6-OMTase activity will be detected in the initial library, clones having other 6-OMTase  
25                  activity may exist. These clones can then be used in further rounds of shuffling to further tailor the 6-OMTase specificity. For example, if activity was detected for erythromycin A, subsequent libraries can be screened first for activity for deserythromycin A, and finally for the deserythromycin oxime. In this way only subtle changes in specificity are expected from each new library.

### Genes and Library Generation

The genes encoding the open reading frames for the midecamycin 3' O-methyltransferase (*mdmC*), the safromycin O-methyltransferase (*safC*), the rapamycin 31-O-methyltransferase (*rapI*), and the FK506 31-O-methyltransferase (*fkbm*) (Figure 13) are  
5 isolated and cloned into an appropriate *E. coli* expression vector (pET22B(+)). These genes, which range from 50-80% identical, are then shuffled by family shuffling to generate a library of genes encoding chimeric O-methyltransferases (OMTase). The library is cloned back into the expression vector and expressed in an appropriate *E. coli* host (BL21(DE3)). This library can now be screened for chimeric enzymes having new properties such as a new  
10 specificity for target methylation.

### Generic Screen for OMTase activity

OMTase activity can be measured in high-throughput by using an assay that measures the transfer of the radiolabeled methyl group of ( $^3\text{H}$ )S-adenosylmethionine to a desired donor molecule (see Figure 15). The assay is based on the transfer of the labeled  
15 methyl group from a highly charged molecule (SAM) to a more hydrophobic molecule (Figure 12). The reaction is extracted with an organic solvent such that unreacted SAM remains in the aqueous phase and the methylated substrate is selectively extracted into the organic phase. The organic phase can then be measured for its content of radioactivity. The advantage of this assay is that it is generally applicable to extractable substrates, it is very  
20 high-through-put, and can be used to screen for activity against a pool of compounds simultaneously. The process is as follows.

*Streptomyces lividans* is a particularly suitable host for at least two reasons. First, it is transformed with high efficiency by plasmid DNA isolated from *E. coli*. Second, it is quite permeable to erythromycin and its analogs, so whole cells rather than lysates can be  
25 assayed. Alternatively, one can use a high throughput format for measuring enzyme activities from *Escherichia coli* or *Bacillus subtilis* cell extracts. Purified enzyme or cell lysate is added to an assay mixture of 50 mM phosphate buffer, pH 7.5, containing 0.4 mM  $\text{MgSO}_4$ , 0.1 mM DTT, 0.1 mM ( $^3\text{H}$ ) S-adenosylmethionine, and 1-10 mM of the target substrate(s). After incubation, the reaction is quenched by extraction with ethylacetate. A  
30 sample (50  $\mu\text{l}$ ) of the organic phase is removed, mixed with scintillant (150  $\mu\text{L}$ ) and measured for radioactivity using a 96 well scintillation counter. Clones from samples having

radioactivity higher than a control sample having no enzyme added is considered positive and can be further investigated in more quantitative assays.

#### Evolution of a Clarithromycin Synthase.

Clarithromycin is 6-O-methyl erythromycin. The current process for the preparation of clarithromycin is a seven step chemical methylation of erythromycin. An enzyme capable of carrying out this chemistry in one step could provide a means of preparing clarithromycin by fermentation or biotransformation (*see* Figure 16). To create such an enzyme, the OMTase library is screened for erythromycin 6-O methylase activity.

The shuffled OMTase library is plated out on solid medium to separate individual clones. Individual colonies are picked into 96 well plates containing LB medium (200  $\mu$ l) and ampicillin (100  $\mu$ g/ml). The plates are grown at 30°C for ten hours or until the cultures have reached an optical density of 0.7. Isopropylthiogalactoside (IPTG) is added to 0.1 mM to induce expression of the MTases, and the cells are incubated for an additional 3 hours. The plates are centrifuged and the supernatant discarded. The cell pellet is resuspended in a lysis buffer (200  $\mu$ l) of 50mM phosphate buffer, pH 7.5, containing 1 mM EDTA, 1 mM DTT, 2  $\mu$ g/ml of polymyxin B sulfate, and 1 mg/ml of T4 lysozyme. The reaction is incubated for 15 minutes at 30°C.

A sample from each well (20  $\mu$ l) is transferred using a 96 head liquid handling station, such as the Multimek™, to a 96 deep well plate containing clarithromycin synthase assay buffer (280  $\mu$ l). The buffer is 50 mM phosphate buffer, pH 7.5, containing 0.4 mM MgSO<sub>4</sub>, 0.1 mM DTT, 0.1 mM (<sup>3</sup>H) S-adenosylmethionine, and 1 mM erythromycin. The reaction is incubated at 30°C for one hour. Ethylacetate (300  $\mu$ L) is added to each well, the plate is shaken vigorously, centrifuged, and a sample (50  $\mu$ L) of the upper organic phase is removed and added to a plate containing scintillant (150  $\mu$ L). The plate is then read using a plate scintillation counter. Any sample having radioactivity in the organic phase higher than that from samples harboring the parental genes or no MTase gene likely contains an enzyme that transfers a methyl group to erythromycin. Since there are five potential hydroxyl groups on erythromycin to which a methyl group might be transferred, it is necessary to discern whether it was transferred to the 6-hydroxyl.

**Secondary assay for Clarithromycin synthase.**

The secondary assay for clarithromycin synthase activity is based on chemical modification with phenyl boronate and analysis by mass spectrometry.

Erythromycin can be O-methylated in five positions, on the 6, 11, or 12 positions of the  
5 macrolide ring, or on either the cladinose or the desosamine moieties. Phenyl boronate bind specifically to cis diols, such as the 11,12 diol of erythromycin. Thus if phenyl boronate binds to the enzymatically methylated erythromycin, the methyl group cannot be located at the 11 or the 12 position. To determine whether the modified erythromycin is clarithromycin the following assay is performed.

10 Enzymatic methylation of erythromycin is performed as described above except the SAM used for the modification is not radiolabeled and the cell extract is from a cell showing a positive radioactivity assay. After extraction from the reaction mixture, the organic phase is analyzed by two dimensional mass spectroscopy (MS/MS), in which the parent ion is fragmented to submolecular fragments (*see* Figure 17).

15 Clarithromycin has a positive ion molecular weight of 748.48, with the positive charge being due to the protonation of the amine of the desosamine moiety. Upon fragmentation of the clarithromycin positive ion, cladinose and the desosamine can be separated from the macrolide ring, however, only molecules containing the desosamine moiety are detected since they carry the amine. Fragmentation of the 748.48 ion results in  
20 two distinctive new ions, 590.4 and 158.12. The 590 ion is 6-O-methyl deserythromycin A (clarithromycin lacking the cladinose moiety). The 158.12 ion is dehydro desosamine, the result of the elimination of the 5-hydroxyl group of the macrolide ring. An MS/MS spectrum of the 748.48 peak having the 590 and the 158 ions is distinctive of erythromycin derivatives methylated on the macrolide ring *i.e.* at the 6, 11, or 12 positions. If the sample shows this  
25 spectrum, then it is further analyzed to determine if it is methylated at the 6 position. The organic extract is treated with an excess of phenylboronate under neutral conditions and then analyzed by mass spectroscopy. Only if the modification is at the 6 position will the 11 and 12 positions be free to form an adduct with the phenylboronate. Thus, the presence of a molecular ion of 834.52, the phenylboronyl adduct of clarithromycin, indicates that the  
30 sample contains clarithromycin and the corresponding clone encodes an erythromycin 6-O-methyltransferase.

It is understood that the examples and embodiments described herein are for illustrative purposes only and that various modifications or changes in light thereof will be suggested to persons skilled in the art and are to be included within the spirit and purview of this application and scope of the appended claims. All publications, patents, and patent applications cited herein are hereby incorporated by reference for all purposes.



**WHAT IS CLAIMED IS:**

- 1                   1. A method for obtaining a library of organic molecule derivatives, the  
2 method comprising contacting an organic molecule with one or more members of a  
3 library of recombinant derivatizing enzymes and other necessary reactants to form the  
4 library of organic molecule derivatives, wherein the derivatizing enzymes catalyze a  
5 reaction selected from the group consisting of:  
6                   a) modification of one or more functional groups present on the  
7 organic molecule;  
8                   b) addition of a chemical moiety onto one or more functional groups  
9 present on the organic molecule; and  
10                  c) introduction of a new functional group.
- 1                   2. The method of claim 1, wherein the method further comprises  
2 contacting the library of organic molecule derivatives with one or more members of a  
3 second library of recombinant derivatizing enzymes and other necessary reactants to form  
4 a further library of organic molecule derivatives, wherein the derivatizing enzymes of the  
5 second library catalyze a reaction selected from the group consisting of:  
6                   a) modification of one or more of the functional groups;  
7                   b) addition of a chemical moiety onto one or more of the functional  
8 groups; and  
9                   c) introduction of a new functional group.
- 1                   3. The method of claim 2, wherein the derivatizing enzymes of the  
2 second library catalyze the modification of, or addition of a chemical moiety onto, a  
3 functional group that was modified or added by the derivatizing enzymes of the first  
4 library.
- 1                   4. The method of claim 1, wherein the one or more members of the  
2 library of organic molecule derivatives is further derivatized by a chemical or enzymatic  
3 reaction after the contacting with the library of recombinant derivatizing enzymes.

1                   5.    The method of claim 1, wherein the library of recombinant  
2    derivatizing enzymes is obtained by a shuffling method.

1                   6.    The method of claim 5, wherein the shuffling method comprises:  
2                        (1) recombining at least first and second forms of a nucleic acid that  
3    encodes a derivatizing enzyme, wherein the first and second forms differ from each other  
4    in two or more nucleotides, to produce a library of recombinant polynucleotides; and  
5                        (2) expressing the library of recombinant polynucleotides to obtain  
6    the library of recombinant derivatizing enzymes.

1                   7.    The method of claim 6, wherein the recombining step is performed in  
2    vitro.

1                   8.    The method of claim 6, wherein the method further comprises:  
2                        (3) recombining at least one recombinant polynucleotide that  
3    encodes a member of the library of recombinant derivatizing enzymes with a further form  
4    of the nucleic acid that encodes a derivatizing enzyme, which is the same or different  
5    from the first and second forms, to produce a further library of recombinant nucleic acids;  
6                        (4) expressing the further library of recombinant polynucleotides to  
7    obtain a further library of recombinant derivatizing enzymes; and  
8                        (5) repeating (3) and (4), as necessary, until the further library of  
9    recombinant derivatizing enzymes contains a desired number of different recombinant  
10   derivatizing enzymes.

1                   9.    The method of claim 8, wherein at least one recombining step is  
2    performed *in vitro*.

1                   10.   The method of claim 5, wherein the shuffling method comprises.  
2                        (1) initiating a polynucleotide amplification process on overlapping  
3    segments of a population of variant polynucleotides under conditions whereby one  
4    segment serves as a template for extension of another segment, to generate a population  
5    of recombinant polynucleotides; and

6 (2) selecting or screening a recombinant polynucleotide for a desired  
7 property.

1 11. The method of claim 10, wherein the overlapping segments are  
2 produced by cleavage of the population of variant polynucleotides.

1 12. The method of claim 11, wherein the cleavage is by DNaseI  
2 digestion.

1 13. The method of claim 10, wherein the overlapping segments are  
2 produced by chemical synthesis.

1 14. The method of claim 10, wherein the overlapping segments are  
2 produced by amplification of the population of polynucleotides.

1 15. The method of claim 10, wherein the population of variant  
2 polynucleotides are allelic variants.

1 16. The method of claim 10, wherein the population of variant  
2 polynucleotides are species variants.

1 17. The method of claim 5, wherein the shuffling method comprises:  
2 (1) hybridizing at least two sets of nucleic acids, wherein a first set  
3 of nucleic acids comprises single-stranded nucleic acid templates and a second set of  
4 nucleic acids comprises at least one set of nucleic acid fragments; and,  
5 (2) elongating, ligating, or both, sequence gaps between the  
6 hybridized nucleic acid fragments, to generate at least substantially full-length chimeric  
7 nucleic acid sequences that correspond to the single-stranded nucleic acid templates,  
8 thereby recombining the set of nucleic acid fragments.

1 18. The method of claim 17, further comprising:  
2 (3) denaturing the at least substantially full-length chimeric nucleic  
3 acid sequences and the single-stranded nucleic acid templates;

4 (4) separating the at least substantially full-length chimeric nucleic  
5 acid sequences from the single-stranded nucleic acid templates by at least one separation  
6 technique; and, fragmenting the separated at least substantially full-length chimeric  
7 nucleic acid sequences by nuclease digestion or physical fragmentation to provide  
8 chimeric nucleic acid fragments.

1 19. The method of claim 1, wherein the organic molecule is a lead  
2 compound.

1 20. The method of claim 1, wherein the organic molecule is a naturally  
2 occurring compound.

1 21. The method of claim 1, wherein the organic molecule is a non-  
2 naturally occurring compound.

1 22. The method of claim 1, wherein the members of the library of  
2 recombinant derivatizing enzymes are contacted with the organic molecule individually.

1 23. The method of claim 1, wherein the members of the library of  
2 recombinant derivatizing enzymes are subdivided into pools prior to contacting the  
3 organic molecule.

1 24. The method of claim 1, wherein the members of the library of  
2 recombinant derivatizing enzymes are contacted with the organic molecule as a mixture  
3 of recombinant derivatizing enzymes.

1 25. The method of claim 1, wherein the recombinant polynucleotides are  
2 expressed by introduction of the recombinant polynucleotides into a replicable genetic  
3 packaging vector so that the encoded recombinant derivatizing enzymes are produced as  
4 fusions with a protein displayed on the surface of a replicable genetic package.

1 26. The method of claim 25, wherein the replicable genetic package is  
2 selected from the group consisting of a bacteriophage, a cell, a spore, and a virus.

- 1                   **27.** The method of claim 1, wherein the derivatizing enzymes catalyze  
2 the modification of one or more functional groups on the organic molecule or the  
3 replacement of one or more of the functional groups with another functional group.
- 1                   **28.** The method of claim 27, wherein the functional group is a hydrogen  
2 and the substitution is by a hydroxyl group.
- 1                   **29.** The method of claim 28, wherein the derivatizing enzyme is selected  
2 from the group consisting of a monooxygenase and a dioxygenase.
- 1                   **30.** The method of claim 27, wherein the derivatizing enzymes catalyze  
2 the introduction of a new functional group onto an organic molecule.
- 1                   **31.** The method of claim 30, wherein the derivatizing enzyme is selected  
2 from the group consisting of a halogenase and a sulfotransferase.
- 1                   **32.** The method of claim 1, wherein the derivatizing enzymes catalyze  
2 the addition of a chemical moiety to one or more of the functional groups.
- 1                   **33.** The method of claim 32, wherein the derivatizing enzymes are  
2 selected from the group consisting of a glycosyltransferase, an acyltransferase, an  
3 amidase, a methyltransferase, and a phosphotransferase.
- 1                   **34.** The method of claim 33, wherein the derivatizing enzyme is a  
2 acyltransferase and the chemical moiety is selected from the group consisting of a vinyl  
3 ester, a trihaloethyl, an ester, a vinyl carbonate, a vinyl carbamate, an oxime ester, an  
4 oxime carbonate, and a bifunctional moiety.
- 1                   **35.** The method of claim 33, wherein the derivatizing enzyme is a  
2 glycosyltransferase and the chemical moiety is selected from the group consisting of a  
3 glycoside, an aminoglycoside, and a glycosidic acid.

1                   36. The method of claim 33, wherein the derivatizing enzyme is a  
2 glycosyltransferase and the organic molecule is selected from the group consisting of  
3 aglycosyl vancomycin HCl, somatostatin, cholic acid, L-thyroxine, nogalamycin,  
4 syringaldazine, aclarubicin, ritodrine HCl, rifamycin, and ristomycin sulfate.

1                   37. The method of claim 33, wherein the derivatizing enzyme is an O-  
2 methyltransferase and the organic molecule is erythromycin.

1                   38. The method of claim 33, wherein the derivatizing enzyme is an  
2 amidase and the chemical moiety is selected from the group consisting of an amide and a  
3 peptide.

1                   39. The method of claim 1, wherein the method further comprises  
2 screening the library of organic molecule derivatives to identify those organic molecule  
3 derivatives that exhibit a desired property.

1                   40. The method of claim 39, wherein the desired property is binding to a  
2 target molecule.

1                   41. The method of claim 40, wherein the target molecule is selected from  
2 the group consisting of a receptor, a signaling protein, and a ligand.

1                   42. The method of claim 39, wherein the method further comprises  
2 screening members of the library of recombinant derivatizing enzymes to identify a  
3 member that catalyzes a modification of the organic molecule that confers upon the  
4 resulting organic molecule derivative the desired property.

1                   43. A method of obtaining an enzyme that catalyzes the synthesis of a  
2 desired organic molecule derivative, the method comprising:  
3                               contacting an organic molecule with members of a library of  
4 recombinant derivatizing enzymes and other necessary reactants to form a library of  
5 organic molecule derivatives;

6 identifying the desired organic molecule derivative in the library of  
7 organic molecule derivatives; and  
8 identifying the member of the library of recombinant derivatizing  
9 enzymes that catalyzes the synthesis of the desired organic molecule derivative.

1 44. The method of claim 43, wherein the members of the library of  
2 recombinant derivatizing enzymes are contacted with the organic molecule individually.

1 45. The method of claim 43, wherein the members of the library of  
2 recombinant derivatizing enzymes are subdivided into pools prior to contacting the  
3 organic molecule.

1 46. A library of recombinant derivatizing enzymes, wherein the  
2 recombinant derivatizing enzymes, when contacted with an organic molecule having one  
3 or more functional groups, catalyze a reaction selected from the group consisting of:  
4 a) modification of one or more of the functional groups;  
5 b) addition of a chemical moiety onto one or more of the functional  
6 groups; and  
7 c) introduction of a new functional group.

1 47. The library of claim 46, wherein the recombinant derivatizing  
2 enzymes each comprise a plurality of blocks of amino acids, which blocks are not  
3 contiguous in a naturally occurring derivatizing enzyme.

1 48. The library of claim 47, wherein the recombinant derivatizing  
2 enzymes each comprise blocks of amino acids that originate from two or more homologs  
3 of the derivatizing enzyme.

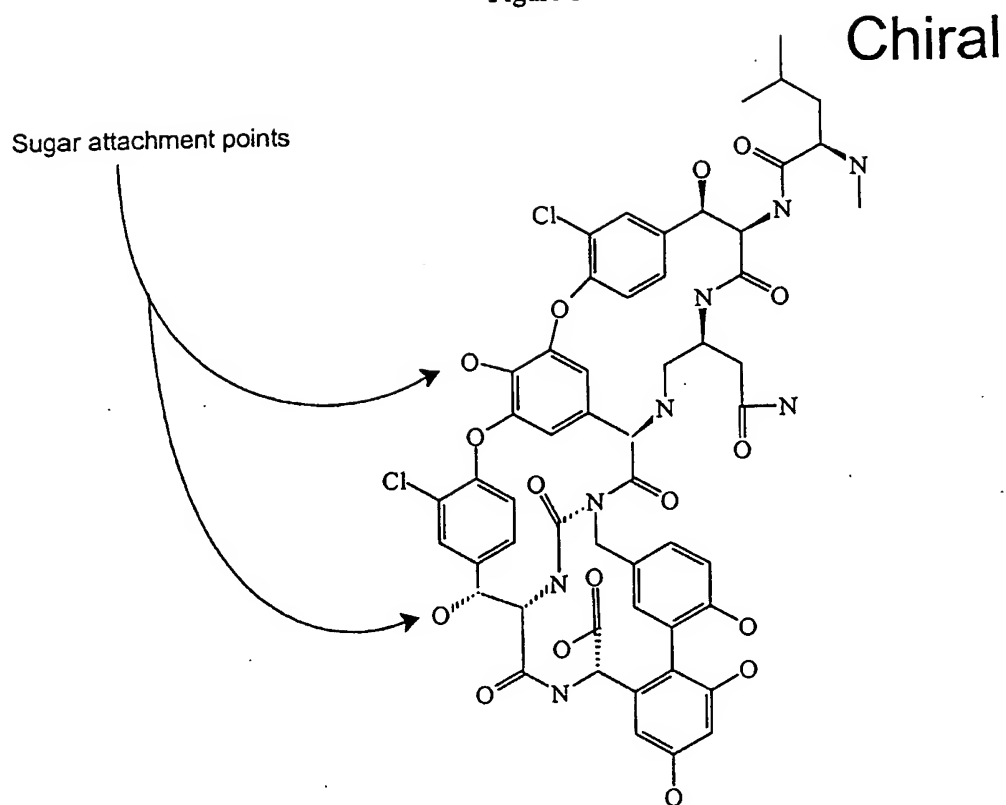
1 49. A library of organic molecule derivatives, wherein the library is  
2 biocatalytically synthesized by contacting an organic molecule having one or more  
3 functional groups with a plurality of members of a library of recombinant derivatizing  
4 enzymes that catalyze a reaction selected from the group consisting of:  
5 a) modification of one or more of the functional groups;

- 6                                   b) addition of a chemical moiety onto one or more of the functional  
7    groups; and  
8                                   c) introduction of a new functional group.

1                           **50.** The library of claim 49, wherein the recombinant derivatizing enzymes  
2    are obtained by:  
3                                   recombining at least first and second forms of a nucleic acid that  
4    encodes a derivatizing enzyme, wherein the first and second forms differ from each other in  
5    two or more nucleotides, to produce a library of recombinant polynucleotides; and  
6                                   expressing the library of recombinant polynucleotides to obtain the  
7    library of recombinant derivatizing enzymes.



Figure 1

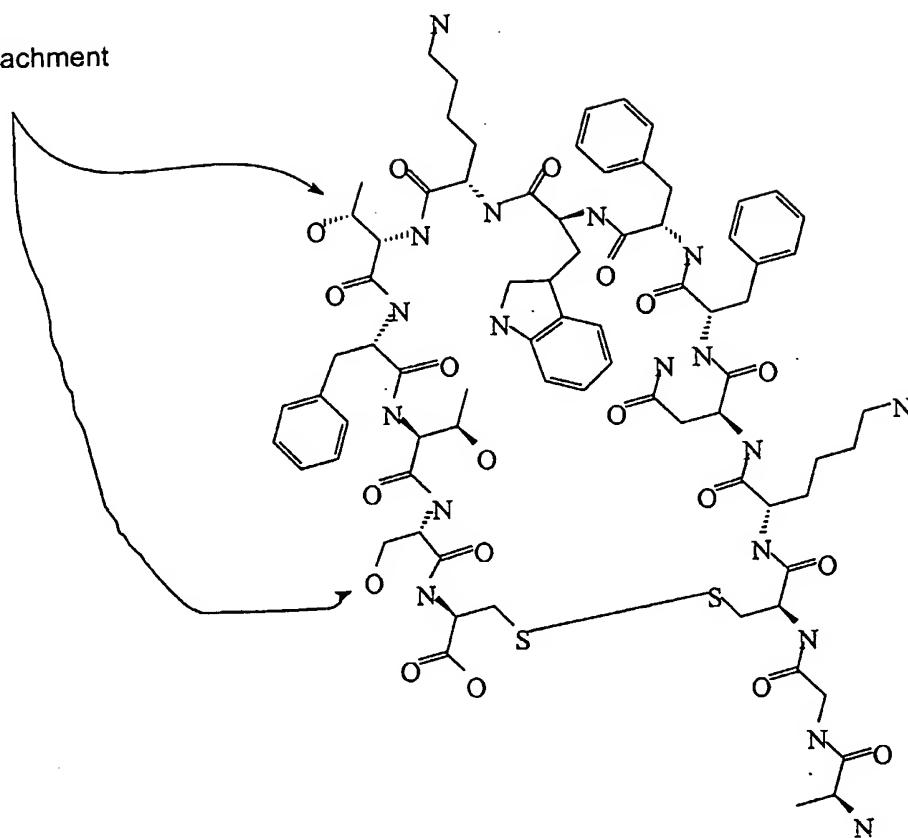


Vancomycin Hydrochloride

Figure 2

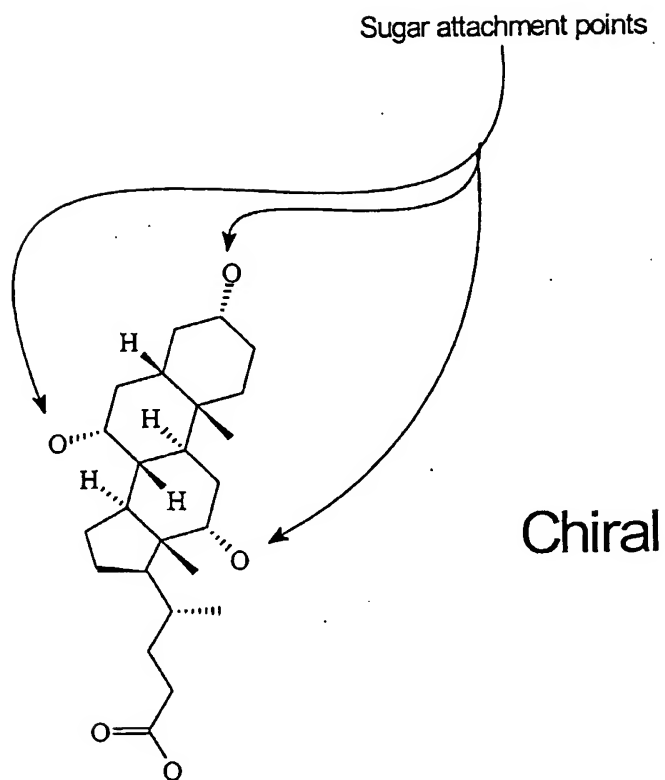
Chiral

Sugar attachment



Somatostatin

Figure 3

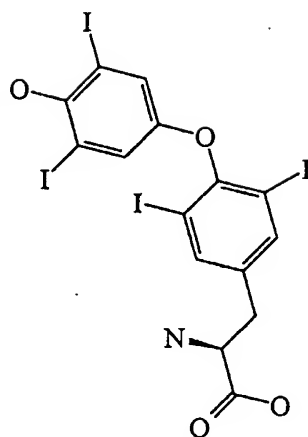


Cholic Acid

Figure 4

Sugar attachment points

Chiral



L-Thyroxine

Figure 5

Chiral

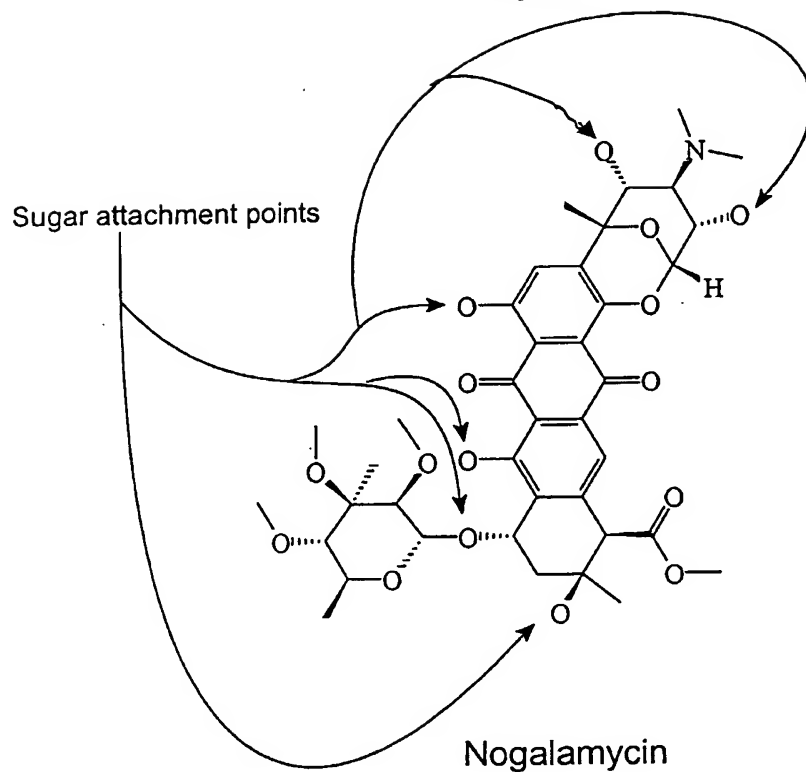
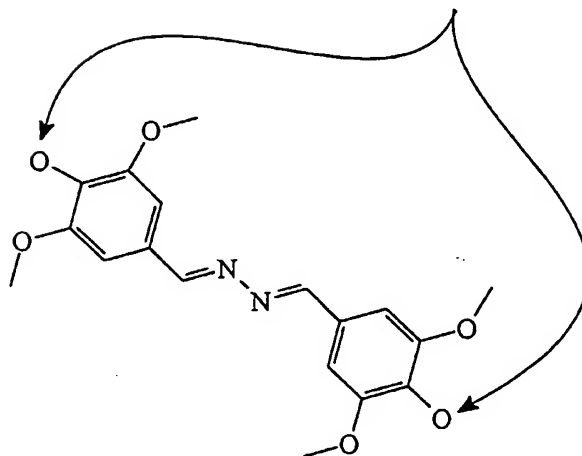


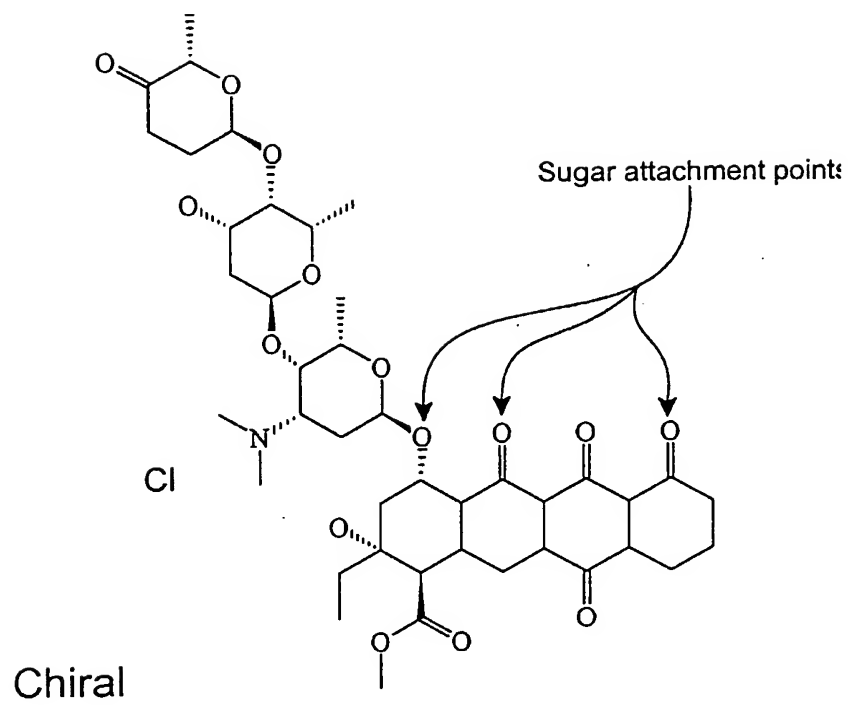
Figure 6

Sugar attachment points



Syringaldazine

Figure 7



Aclarubicin

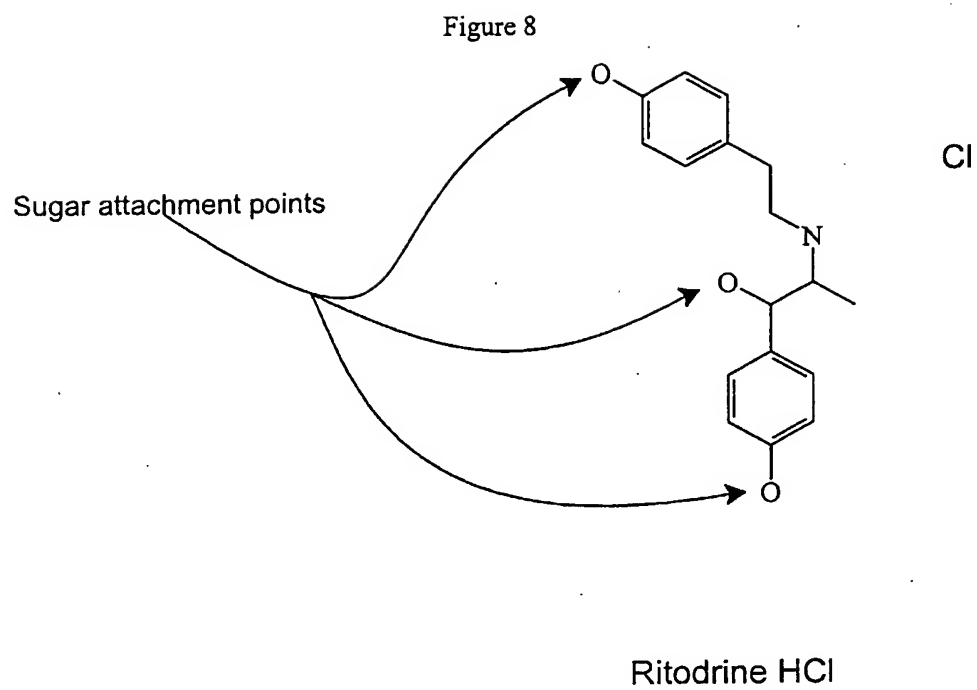
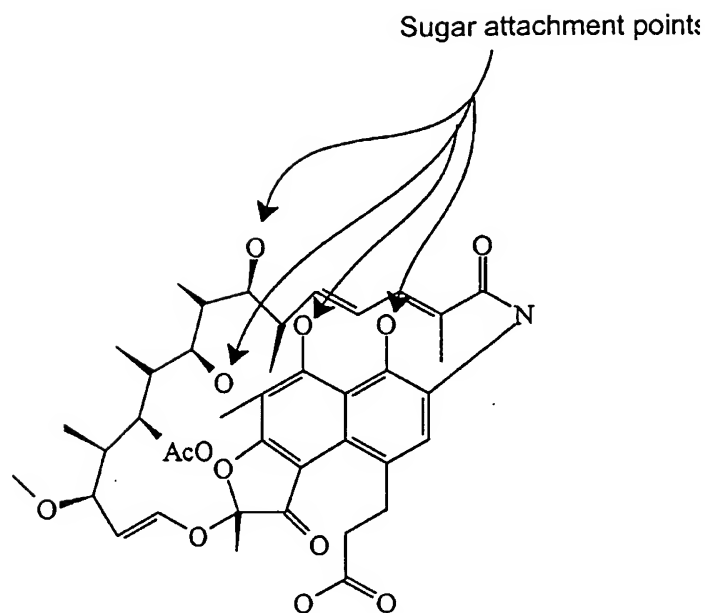




Figure 9



Rifamycin

Figure 10

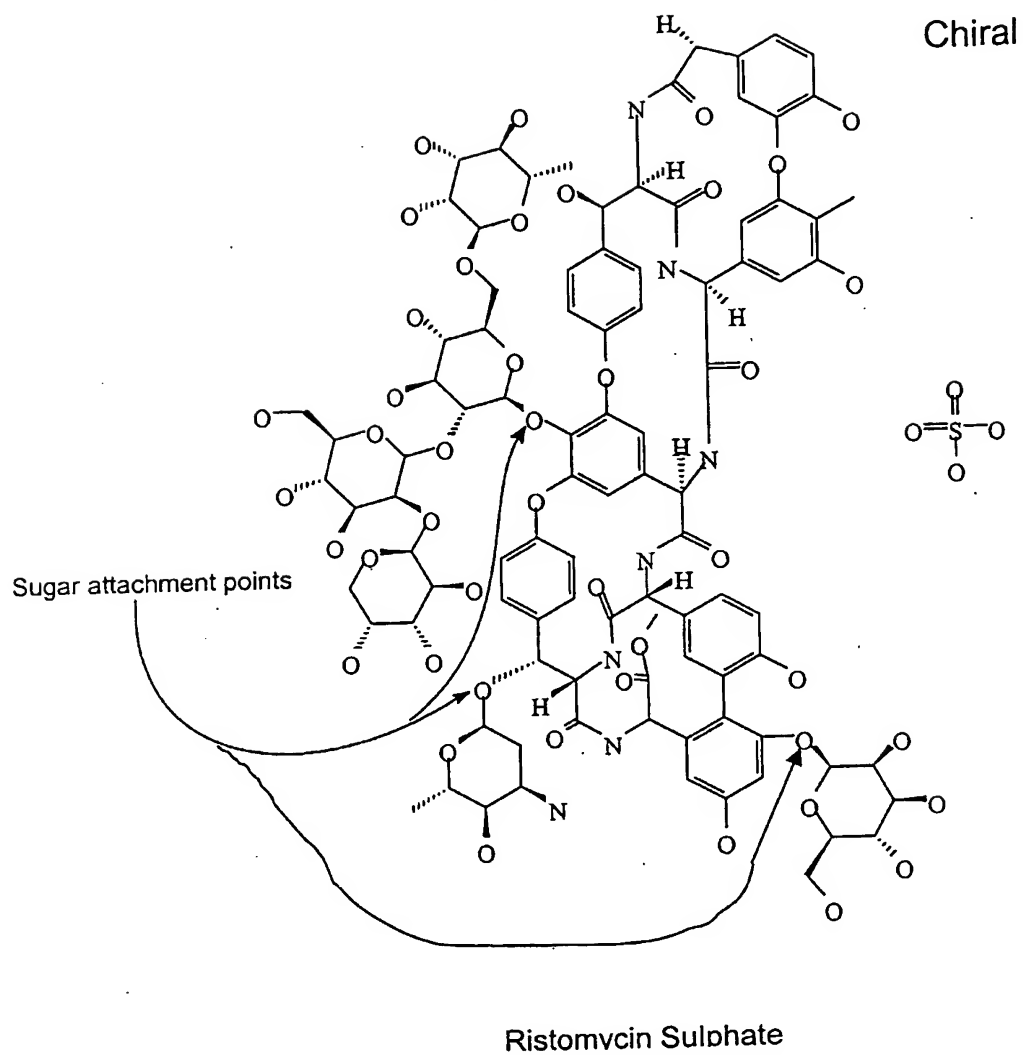


Figure 11

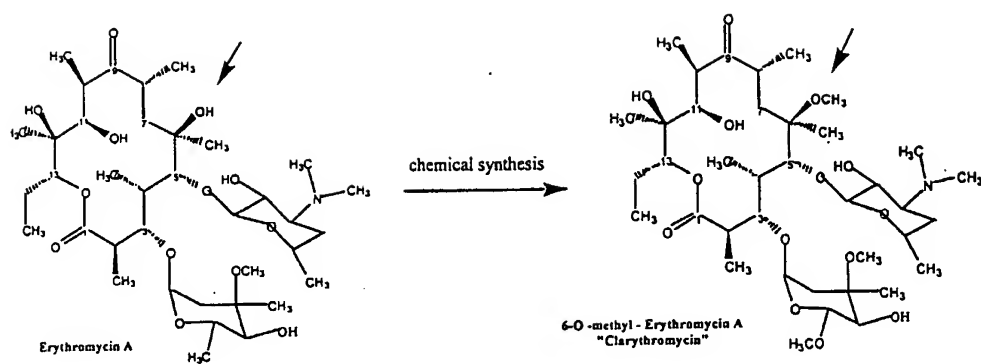


Figure 12

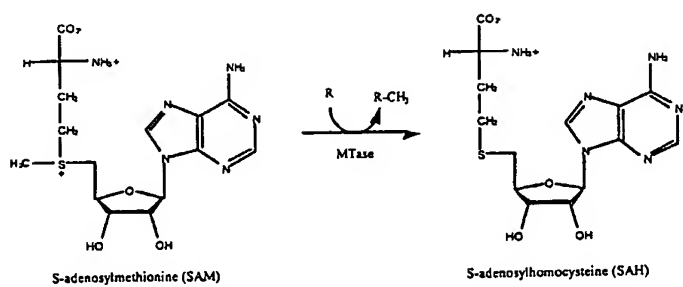


Figure 13

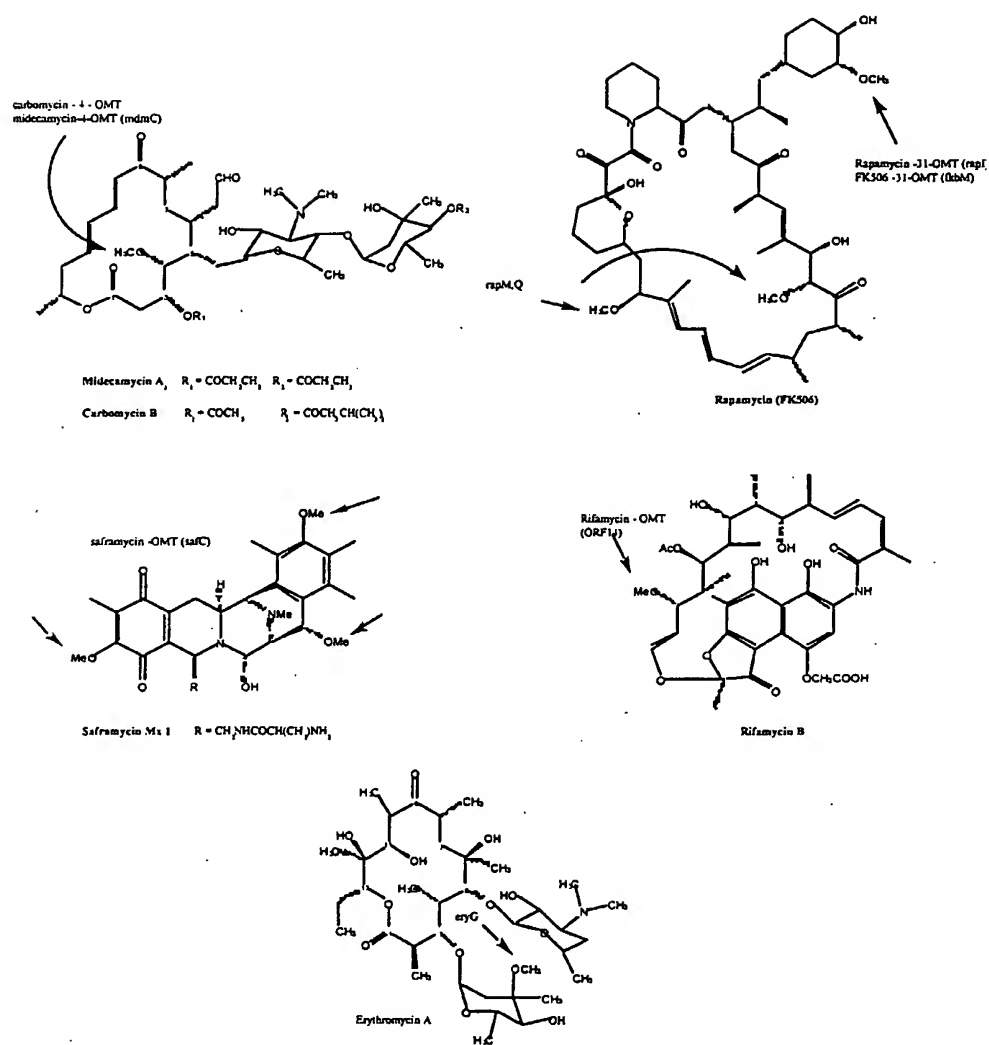


Figure 14

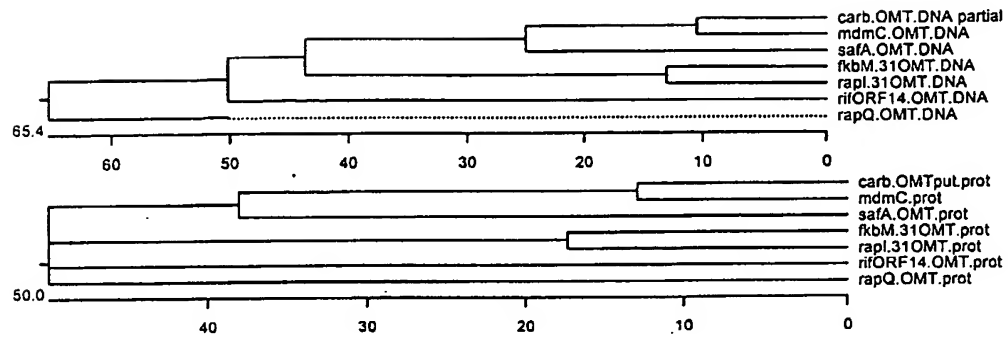
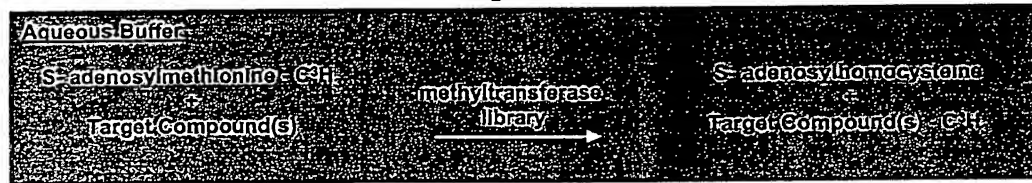
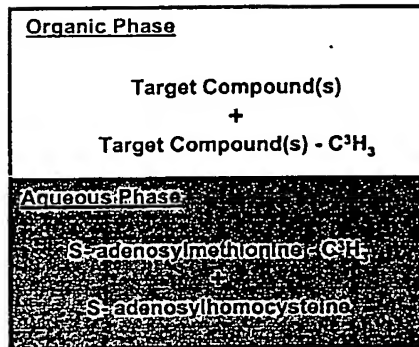


Figure 15



## Organic Extraction



→ Screen for presence of radioactivity in organic phase

↓  
Analyze positives by stringent secondary assay

Figure 16

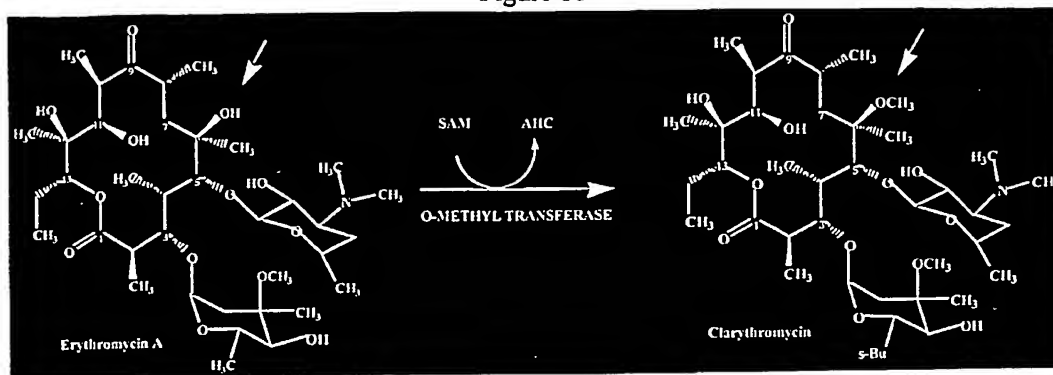


Figure 17

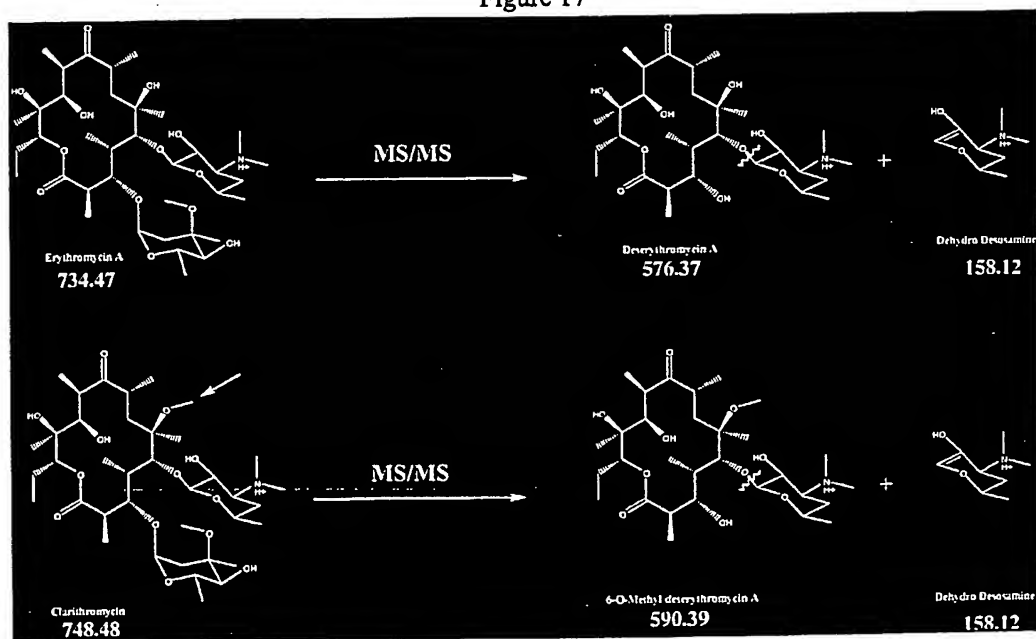


Figure 18

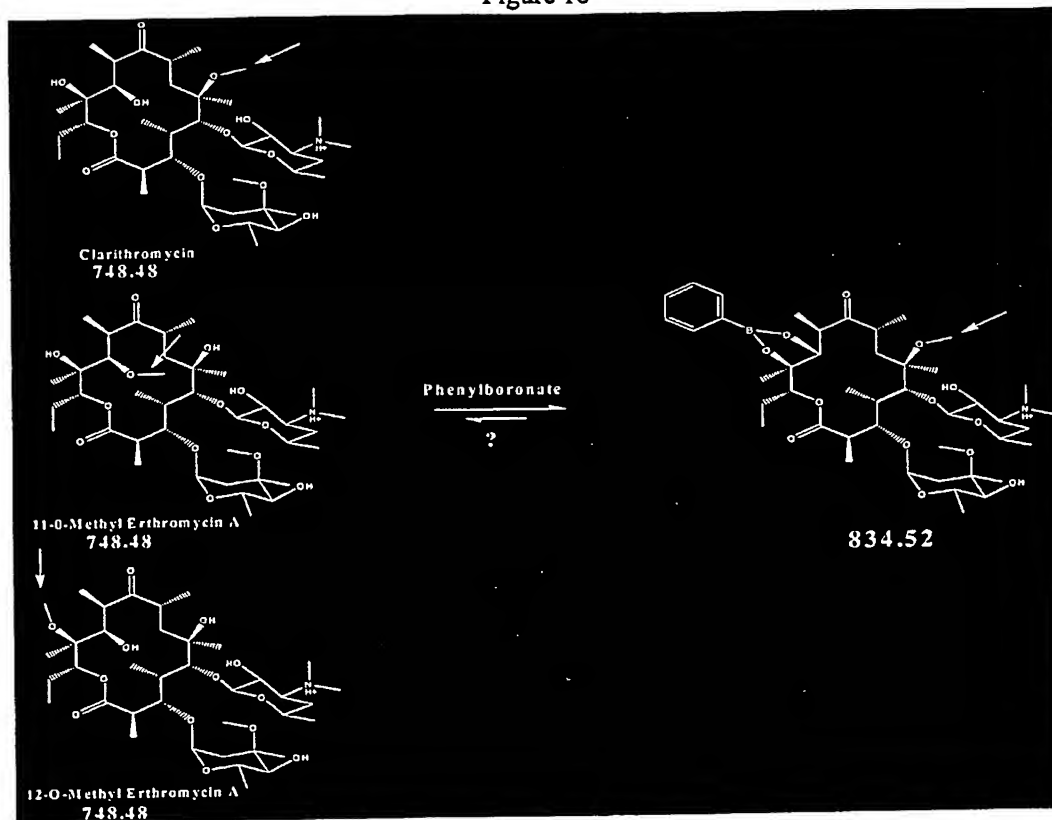
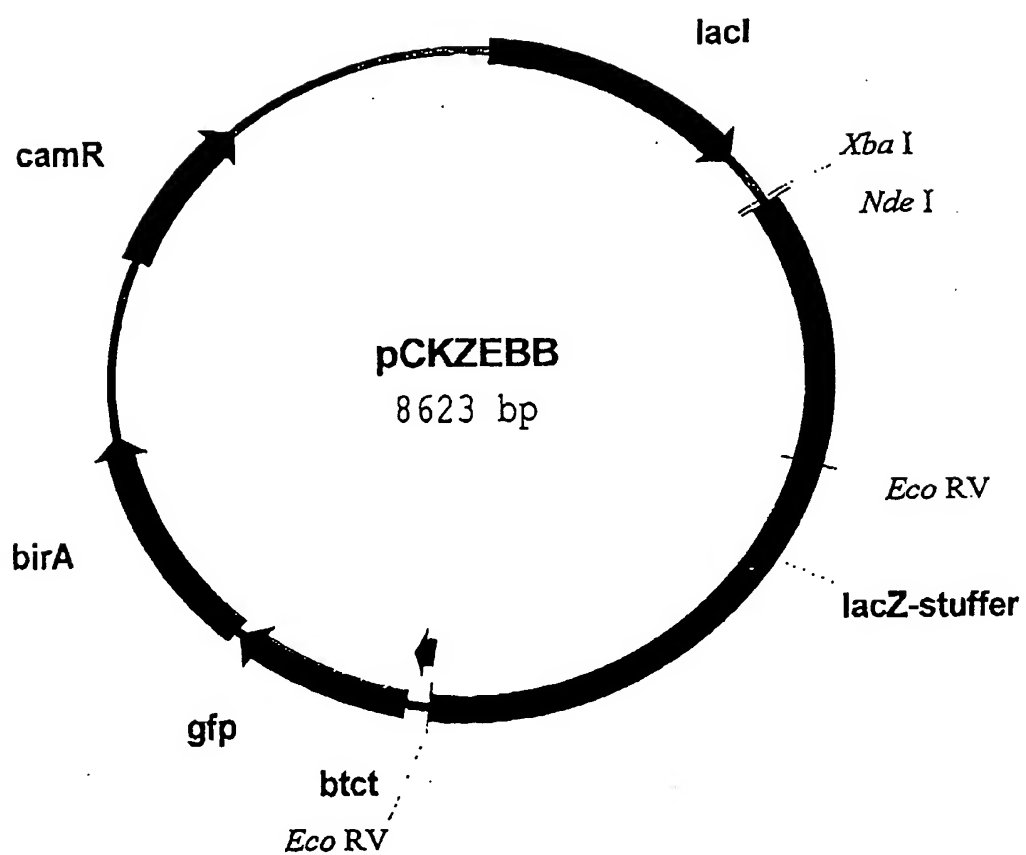


Figure 19





**Sequence Listing**

SEQ ID NO:19: Sequence of pCKZEBB vector

ACCCGACACCATCGAATGGCGCAAAACCTTTCGCGGTATGGCATGATAGCGCCC  
GGAAGAGAGTCAATTCAGGGTGGTGAATGTGAAACCAGTAACGTTATACGATGT  
CGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCGTGGTGAACCAGGC  
CAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGC  
TGAATTACATTCCCAACCGCGTGGCACAACAACCTGGCGGGCAAACAGTCGTTGC  
TGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAATTGTTCG  
GGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTTCGATGGT  
AGAACGAAGCGGCGTCGAAGCCTGTAAAGCGGCGGTGCACAATCTTCTCGCGCA  
ACGCGTCAGTGGGCTGATCATTAACCTATCCGCTGGATGACCAGGATGCCATTGC  
TGTGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAG  
ACACCCATCAACAGTATTATTTTCTCCCATGAAGACGGTACGCGACTGGGCGTG  
GAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATT  
AGTTCTGTCTCGGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCA  
ATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGTGCCATGTCCGGTT  
TTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGG  
TTGCCAACGATCAGATGGCGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGC  
TGCGCGTTGGTGCGGACATCTCGGTAGTGGGATACGACGATACCGAAGACAGCT  
CATGTTATATCCCGCCGTTAACCACCATCAAACAGGATTTTCGCTGCTGGGGCA  
AACCAGCGTGGACCGCTTGCTGCAACTCTCTCAGGGCCAGGCGGTGAAGGGCAA  
TCAGCTGTTGCCCCTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATAC  
GCAAACCGCCTCTCCCCGCGGTTGGCCGATTCATTAATGCAGCTGGCACGACA  
GGTTTCCCGACTGGAAAGCGGGCAGTGAGCGGTACCCGATAAAAGCGGCTTCCT  
GACAGGAGGCCGTTTTGTTTTGCAGCCCACCTCAACGCAATTAATGTGAGTTAG  
CTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGT  
GTGGAATTGTGAGCGGATAACAATTTACACAGGAAACAGCTATGACCATGATT  
ACGAATTTCTAGAGAAGGAGATATACATATGACCATGATTACGGATTCAGTGGC  
CGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACCTAATCGC  
CTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGCCCGCACC

## Sequence Listing (cont'd)

GATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTGCCTGG  
TTTCCGGCACCAGAAGCGGTGCCGAAAGCTGGCTGGAGTGCATCTTCCTGAG  
GCCGATACTGTCGTCGTCCCCTCAAACCTGGCAGATGCACGGTTACGATGCGCCC  
ATCTACACCAACGTAACCTATCCATTACGGTCAATCCGCCGTTTGTTCACCG  
AGAATCCGACGGGTTGTTACTCGCTCACATTTAATGTTGATGAAAGCTGGCTAC  
AGGAAGGCCAGACGCGAATTATTTTGTATGGCGTTAACTCGGCGTTTCATCTGTG  
GTGCAACGGGGCGCTGGGTTCGGTTACGGCCAGGACAGTCGTTTGCCGTCTGAATT  
TGACCTGAGCGCATTTTTACGCGCCGGAGAAAACCGCCTCGCGGTGATGGTGCT  
GCGTTGGAGTGACGGCAGTTATCTGGAAGATCAGGATATGTGGCGGATGAGCGG  
CATTTTCCGTGACGTCTCGTTGCTGCATAAACCGACTACACAAATCAGCGATTTC  
CATGTTGCCACTCGCTTTAATGATGATTTACGCCGCGCTGTACTGGAGGCTGAAG  
TTCAGATGTGCGGCGAGTTGCGTGACTACCTACGGGTAACAGTTTCTTTATGGCA  
GGGTGAAACGCAGGTCGCCAGCGGCACCGCGCCTTTCGGCGGTGAAATTATCGA  
TGAGCGTGGTGGTTATGCCGATCGCGTCACACTACGTCTGAACGTCGAAAACCC  
GAAACTGTGGAGCGCCGAAATCCCGAATCTCTATCGTGCGGTGGTTGAACTGCA  
CACCGCCGACGGCACGCTGATTGAAGCAGAAGCCTGCGATGTCGGTTTCCGCGA  
GGTGCGGATTGAAAATGGTCTGCTGCTGCTGAACGGCAAGCCGTTGCTGATTCTG  
AGGCGTTAACCGTCACGAGCATATCCTCTGCATGGTCAGGTCATGGATGAGCA  
GACGATGGTGCAGGATATCTTTTTGACACCAGACCAACTGGTAATGGTAGCGAC  
CGGCGCTCAGCTGGAATTCCGCCGATACTGACGGGCTCCAGGAGTCGTGCCAC  
CAATCCCCATGTGGAAACCGTCGATATTCAGCCATGTGCCTTCTTCGCGTGCGAG  
CAGATGGCGATGGCTGGTTTCCATCAGTTGCTGTTGACTGTAGCGGCTGATGTTG  
AACTGGAAGTCGCCGCGCCACTGGTGTGGGCCATAATTCAATTCGCGCGTCCCG  
CAGCGCAGACCGTTTTTCGCTCGGGAAGACGTACGGGGTATACATGTCTGACAAT  
GGCAGATCCCAGCGGTCAAAACAGGCGGCAGTAAGGCGGTGCGGATAGTTTTCT  
TGCGGCCCTAATCCGAGCCAGTTTACCCGCTCTGCTACCTGCGCCAGCTGGCAGT  
TCAGGCCAATCCGCGCCGGATGCGGTGTATCGCTCGCCACTTCAACATCAACGG  
TAATCGCCATTTGACCACTACCATCAATCCGGTAGGTTTTCCGGCTGATAAATAA  
GGTTTTCCCTGATGCTGCCACGCGTGAGCGGTGTAATCAGCACCGCATCAGC  
AAGTGTATCTGCCGTGCACTGCAACAACGCTGCTTCGGCCTGGTAATGGCCCCG

## Sequence Listing (cont'd)

CGCCTTCCAGCGTTCGACCCAGGCGTTAGGGTCAATGCGGGTCGCTTCACTTACG  
CCAATGTCGTTATCCAGCGGTGCACGGGTGAACTGATCGCGCAGCGGCGTCAGC  
AGTTGTTTTTTATCGCCAATCCACATCTGTGAAAGAAAGCCTGACTGGCGGTAA  
ATTGCCAACGCTTATTACCCAGCTCGATGCAAAAATCCATTCGCTGGTGGTCAG  
ATGCGGGATGGCGTGGGACGCGGCGGGGAGCGTCACACTGAGGTTTTCCGCCAG  
ACGCCACTGCTGCCAGGCGCTGATGTGCCCGGCTTCTGACCATGCGGTCGCGTTC  
GGTTGCACTACGCGTACTGTGAGCCAGAGTTGCCCGGCGCTCTCCGGCTGCGGT  
AGTTCAGGCAGTTCAATCAACTGTTTACCTTGTGGAGCGACATCCAGAGGCACT  
TCACCGCTTGCCAGCGGCTTACCATCCAGCGCCACCATCCAGTGCAGGAGCTCG  
TTATCGCTATGACGGAACAGGTATTCGCTGGTCACTTCGATGGTTTTGCCCGGATA  
AACGGAAGTGGAAAACTGCTGCTGGTGTGTTTTGCTTCCGTCAGCGCTGGATGCG  
GCGTGCGGTCGGCAAAGACCAGACCGTTCATACAGAACTGGCGATCGTTCGGCG  
TATCGCCAAAATCACCGCCGTAAGCCGACCACGGGTTGCCGTTTTTCATCATATTT  
AATCAGCGACTGATCCACCCAGTCCCAGACGAAGCCGCCCTGTAAACGGGGATA  
CTGACGAAACGCCTGCCAGTATTTAGCGAAACCGCCAAGACTGTTACCCATCGC  
GTGGGCGTATTCGCAAAGGATCAGCGGGCGCGTCTCTCCAGGTAGCGAAAGCCA  
TTTTTTGATGGACCATTTTCGGCACAGCCGGAAGGGCTGGTCTTCATCCACGCGC  
GCGTACATCGGGCAAATAATATCGGTGGCCGTGGTGTGCGCTCCGCCGCCTTCA  
TACTGCACCGGGCGGGAAGGATCGACAGATTTGATCCAGCGATACAGCGCGTGC  
TGATTAGCGCCGTGGCCTGATTCATTCCCCAGCGACCAGATGATCACACTCGGG  
TGATTACGATCGCGCTGCACCATTCGCGTTACGCGTTCGCTCATCGCCGGTAGCC  
AGCGCGGATCATCGGTCAGACGATTCATTGGCACCATGCCGTGGGTTTCAATAT  
TGGCTTCATCCACCACATACAGGCCGTAGCGGTCGCACAGCGTGTACCACAGCG  
GATGGTTCGGATAATGCGAACAGCGCACGGCGTTAAAGTTGTTCTGCTTCATCA  
GCAGGATATCGAAGGTGGTGGTTCTGCTCAGCGTCTGTTCCACATCCTGGACGCT  
CAGAAAATCGAATGGCACGGTCCGAAAGCACGAGGGGAAATCTGATGGCTAGC  
AAAGGAGAAGAACTTTTCACTGGAGTTGTCCCAATTCTTGTTGAATTAGATGGT  
GATGTTAATGGGCACAAATTTTCTGTGAGTGGAGAGGGTGAAGGTGATGCTACA  
TACGGAAAACCTTACCCTTAAATTTATTTGCACTACTGGAAAACCTACCTGTTCCAT  
GGCCAACACTTGTCATACTTTCTCTTATGGTGTTCATGCTTTTCCCGTTATCCG

## Sequence Listing (cont'd)

GATCAGATGAAACGGCATGACTTTTTCAAGAGTGCCATGCCCCGAAGGTTATGTA  
CAGGAACGCACTATATCTTTCAAAGATGACGGGAACTACAAGACGCGTGCTGAA  
GTCAAGTTTGAAGGTGATACCCTTGTTAATCGTATCGAGTTAAAAGGTATTGATT  
TTAAAGAAGATGGAAACATTCTCGGACACAACTCGAGTACAACATACTCAC  
ACAATGTATACATCACGGCAGACAAACAAAAGAATGGAATCAAAGCTAACTTC  
AAAATTCGCCACAACATTGAAGATGGATCCGTTCAACTAGCAGACCATTATCAA  
CAAAATACTCCAATTGGCGATGGCCCTGTCCTTTTACCAGACAACCATTACCTGT  
CGACACAATCTGCCCTTTCGAAAGATCCCAACGAAAAGCGTGACCACATGGTCC  
TTCTTGAGTTTGTAAGTCTGCTGGGATTACACATGGCATGGATGAGCTCTACAA  
ATAAGGAGACAATTTTCATGAAGGATAACACCGTGCCACTGAAATTGATTGCCCT  
GTTAGCGAACGGTGAATTTCACTCTGGCGAGCAGTTGGGTGAAACGCTGGGAAT  
GAGCCGGGCGGCTATTAATAAACACATTCAGACACTGCGTGACTGGGGCGTTGA  
TGTCTTTACCGTTCCGGGTAAAGGATACAGCCTGCCTGAGCCTATCCAGTTACTT  
AATGCTAAACAGATATTGGGTCAGCTGGATGGCGGTAGTGTAGCCGTGCTGCCA  
GTGATTGACTCCACGAATCAGTACCTTCTTGATCGTATCGGAGAGCTTAAATCGG  
GCGATGCTTGCATTGCAGAATACCAGCAGGCTGGCCGTGGTCGCCGGGGTCGGA  
AATGGTTTTTCGCCTTTTGGCGCAAACCTTATATTTGTCGATGTTCTGGCGTCTGGA  
ACAAGGCCCGGCGGCGGCGATTGGTTTAAGTCTGGTTATCGGTATCGTGATGGC  
GGAAGTATTACGCAAGCTGGGTGCAGATAAAGTTCGTGTTAAATGGCCTAATGA  
CCTCTATCTGCAGGATCGCAAGCTGGCAGGCATTCTGGTGGAGCTGACTGGCAA  
AACTGGCGATGCGGCGCAAATAGTCATTGGAGCCGGGATCAACATGGCAATGC  
GCCGTGTTGAAGAGAGTGTCGTTAATCAGGGGTGGATCACGCTGCAGGAAGCGG  
GGATCAATCTCGATCGTAATACGTTGGCGGCCATGCTAATACGTGAATTACGTG  
CTGCGTTGGAACCTTTCGAACAAGAAGGATTGGCACCTTATCTGTCGCGCTGGG  
AAAAGCTGGATAATTTTATTAATCGCCCAGTGAACTTATCATTGGTGATAAAG  
AAATATTTGGCATTTCACGCGGAATAGACAAACAGGGGGCTTTATTACTTGAGC  
AGGATGGAATAATAAAACCCTGGATGGGCGGTGAAATATCCCTGCGTAGTGAG  
AAAAATAAGCCCGGGCAAGCTTGACCTGTGAAGTGAAAAATGGCGCACATTGT  
GCGACATTTTTTTTGTCTGCCGTTTACCGCTACTGCGTCACGGATCCCCACGCG  
CCTGTAGCGGCGCATTAAGCGCGCGGGGTGTGGTGGTTACGCGCAGCGTGACCG

## Sequence Listing (cont'd)

CTACACTTGCCAGCGCCCTAGCGCCCGCTCCTTTCGCTTTCTTCCCTTCCTTTCTC  
GCCACGTTTCGCCGGCTTTCCCCGTCAGCTCTAAATCGGGGCATCCCTTTAGGGT  
TCCGATTTAGTGCTTTACGGCACCTCGACCCCAAAAACTTGATTAGGGTGATG  
GTTACGTAAGTGGGCCATCGCCCTGATAGACGGTTTTTCGCCCTTGACGTTGGA  
GTCCACGTTCTTTAATAGTGGACTCTTGTTCCAACTGGAACAACACTCAACCCT  
ATCTCGGTCTATTCTTTTGATTTATAAGGGATTTTGCCGATTCGGCCTATTGGTT  
AAAAAATGAGCTGATTTAACAAAAATTTAACGCGAATTTTAACAAAATATTAAC  
GTTTACAATTTTCAGGTGGCACTTTTCGGGGAAATGTGCGCGGAACCCCTATTTGT  
TTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGTCGAGACGTTGGGTGA  
GGTTCCAACCTTTCACCATAATGAAATAAGATCACTACCGGGCGTATTTTTTGAGT  
TATCGAGATTTTCAGGAGCTAAGGAAGCTAAAATGGAGAAAAAAATCACTGGA  
TATACCACCGTTGATATATCCCAATGGCATCGTAAAGAACATTTTGAGGCATTTT  
AGTCAGTTGCTCAATGTACCTATAACCAGACCGTTCAGCTGGATATTACGGCCTT  
TTTAAAGACCGTAAAGAAAAATAAGCACAAAGTTTTATCCGGCCTTTATTACATT  
CTTGCCCGCCTGATGAATGCTCATCCGGAGTTCCGTATGGCAATGAAAGACGGT  
GAGCTGGTGATATGGGATAGTGTTACCCCTTGTTACACCGTTTTCCATGAGCAA  
CTGAAACGTTTTTCATCGCTCTGGAGTGAATACCACGACGATTTCCGGCAGTTTCT  
ACACATATATTCGCAAGATGTGGCGTGTTACGGTGAAAACCTGGCCTATTTCCCT  
AAAGGGTTTTATTGAGAATATGTTTTTCGTCTCAGCCAATCCCTGGGTGAGTTTCA  
CCAGTTTTGATTTAAACGTGGCCAATATGGACAACCTTCTTCGCCCCCGTTTTCAC  
CATGGGCAAATATTATACGCAAGGCGACAAGGTGCTGATGCCGCTGGCGATTCA  
GGTTCATCATGCCGTCTGTGATGGCTTCCATGTCCGCAGAATGCTTAATGAATTA  
CAACAGTACTGCGATGAGTGGCAGGGCGGGGCGTAATTTTTTTAAGGCAGTTAT  
TGGTGCCCTTAAACGCCTGGTGCTACGCCTGAATAAGTGATAATAAGCGGATGA  
ATGGCAGAAATTCGAAAGCAAATTCGACCCGGTCGTCGGTTCAGGGCAGGGTTCG  
TTAAATAGCCGCTTATGTCTATTGCTGGTTTACCGGTTTATTGACTACCGGAAGC  
AGTGTGACCGTGTGCTTCTCAAATGCCTGAGGCCAGTTTGCTCAGGCTCTCCCCG  
TGGAGGTAATAATTGCTCGACATGACCAAAATCCCTTAACGTGAGTTTTCGTTCC  
ACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTT  
TCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGT

## Sequence Listing (cont'd)

TTGTTTGCCGGATCAAGAGCTACCAACTCTTTTCCGAAGGTAAGTGGCTTCAGC  
AGAGCGCAGATACCAAATACTGTCCTTCTAGTGTAGCCGTAGTTAGGCCACCAC  
TTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAG  
TGGCTGCTGCCAGTGGCGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGAT  
AGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAG  
CCCAGCTTGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTA  
TGAGAAAGCGCCACGCTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAG  
CGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCT  
GGTATCTTTATAGTCCTGTGCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTG  
TGATGCTCGTCAGGGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTT  
TTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATG

## INTERNATIONAL SEARCH REPORT

Intern al Classification No

PCT/US 00/22080

A. CLASSIFICATION OF SUBJECT MATTER  
IPC 7 C12N15/52 C12P1/00 C12P19/62

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 7 C12N C12P

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, BIOSIS, CHEM ABS Data, MEDLINE

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	<p>FRED C. CHRISTIANS ET AL.: "Directed evolution of thymidine kinase for AZT phosphorylation using DNA family shuffling" NATURE BIOTECHNOLOGY., vol. 17, no. 3, March 1999 (1999-03), pages 259-264, XP002154178 NATURE PUBLISHING., US ISSN: 1087-0156 cited in the application abstract page 259, left-hand column, paragraph 1 -page 261, left-hand column, paragraph 1 page 263, right-hand column, paragraph 2 -page 264, left-hand column, paragraph 2 --- -/--</p>	43-50

☒ Further documents are listed in the continuation of box C.☒ Patent family members are listed in annex.

## \* Special categories of cited documents:

- \*A\* document defining the general state of the art which is not considered to be of particular relevance
- \*E\* earlier document but published on or after the international filing date
- \*L\* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- \*O\* document referring to an oral disclosure, use, exhibition or other means
- \*P\* document published prior to the international filing date but later than the priority date claimed

\*T\* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

\*X\* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

\*Y\* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

\*G\* document member of the same patent family

Date of the actual completion of the international search

6 December 2000

Date of mailing of the international search report

22/12/2000

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Montero Lopez, B

## INTERNATIONAL SEARCH REPORT

Intern al Patent No

PCT/US 00/22080

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	ZHANG J-H ET AL: "Directed evolution of a fucosidase from galactosidase by DNA shuffling and screening" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA,US,NATIONAL ACADEMY OF SCIENCE. WASHINGTON, vol. 94, 1 April 1997 (1997-04-01), pages 4504-4509, XP002087464 ISSN: 0027-8424 the whole document	43-50
X	HANSSON L O ET AL: "Evolution of differential substrate specificities in Mu class glutathione transferases probed by DNA shuffling." JOURNAL OF MOLECULAR BIOLOGY, (1999 MAR 26) 287 (2) 265-76. , XP002154179 page 272, left-hand column, paragraph 3 -right-hand column, paragraph 3 page 267, left-hand column, paragraph 2 -page 271, left-hand column, last paragraph abstract	43-48
X	MOZHAEV V V ET AL: "Regioselective Enzymatic Acylation as a Tool for Producing Solution-Phase Combinatorial Libraries" TETRAHEDRON,NL,ELSEVIER SCIENCE PUBLISHERS, AMSTERDAM, vol. 54, no. 16, 16 April 1998 (1998-04-16), pages 3971-3982, XP004162174 ISSN: 0040-4020	49,50
Y	page 3973, last paragraph; figure 2 page 3978, paragraph 2	1-45
Y	KUCHNER O ET AL: "Directed evolution of enzyme catalysts" TRENDS IN BIOTECHNOLOGY,GB,ELSEVIER PUBLICATIONS, CAMBRIDGE, vol. 15, no. 12, 1 December 1997 (1997-12-01), pages 523-530, XP004097432 ISSN: 0167-7799 the whole document	1-45
P,X	WO 00 09682 A (MAXYGEN, INC.) 24 February 2000 (2000-02-24) page 3, line 11 -page 7, line 4 page 17, line 25 -page 90, line 28	1-50



INTERNATIONAL SEARCH REPORT  
Information on patent family members

International Publication No  
PCT/US 00/22080

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0009682 A	24-02-2000	AU 5347999 A	06-03-2000

# INFORMATION SHEET

<b>Application For</b>	PATENT
<b>Country</b>	U.S.A
<b>Title</b>	Stabilized Biocatalysts and Methods of Bioconversion Using the Same
<b>Assignee</b>	
Name	GENOFOCUS CO., LTD.
Address	A Daeduk biocommunity, 461-6, Jeonmin-dong, Yuseung-gu, Daejeon-city 305-390, Republic of Korea
Representative	PAN, Jae-Gu
Nationality	Republic of Korea
<b>Inventor</b>	
Name	KWON, Seok-Joon
Address	A Daeduk biocommunity, 461-6, Jeonmin-dong, Yuseung-gu, Daejeon-city 305-390, Republic of Korea
Nationality	Republic of Korea
<b>Inventor</b>	
Name	CHOI, Soo-Keun
Address	A Daeduk biocommunity, 461-6, Jeonmin-dong, Yuseung-gu, Daejeon-city 305-390, Republic of Korea
Nationality	Republic of Korea
<b>Inventor</b>	
Name	JUNG, Heung-Chae
Address	A Daeduk biocommunity, 461-6, Jeonmin-dong, Yuseung-gu, Daejeon-city 305-390, Republic of Korea
Nationality	Republic of Korea
<b>Inventor</b>	
Name	PAN, Jae-Gu
Address	A Daeduk biocommunity, 461-6, Jeonmin-dong, Yuseung-gu, Daejeon-city 305-390, Republic of Korea
Nationality	Republic of Korea

<b>Inventor</b>	
Name	CHANG, Joon-Sung
Address	A Daeduk biocommunity, 461-6, Jeonmin-dong, Yusung-gu, Daejon-city 305-390, Republic of Korea
Nationality	Republic of Korea
<b>Inventor</b>	
Name	PARK, Tae-Jung
Address	A Daeduk biocommunity, 461-6, Jeonmin-dong, Yusung-gu, Daejon-city 305-390, Republic of Korea
Nationality	Republic of Korea
<b>International Application No.</b>	PCT/KR02/00671 (April 9, 2002)
<b>International Publication No.</b>	WO 03/102203
<b>Filing Deadline</b>	By October 9, 2004